AD-A270 083

N PAGE

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE

3. REPORT TYPE AND DATES COVERED
FINAL/01 OCT 90 TO 30 SEP 92

4. TITLE AND SUBTITLE
SPECTRAL ANALYSIS ON THE CANONICAL AUTOREGRESSIVE DECOMPOSITION (U)

5. FUNDING NUMBERS
2304/A6/ES
AFOSR-91-0024

6. AUTHOR(S)
Professor Fadil Santosa

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Delaware
Dept of Mathematical Sciences
Newark, DE 19716

8. PERFORMING ORGANIZATION REPORT NUMBER
AEOSR-TR- 93 0 05

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NM
110 DUNCAN AVE, SUITE B115
BOLLING AFB DC 20332-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFOSR-91-0024

DTIC ELECTE OCT 05 1993 S B D

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

12b. DISTRIBUTION CODE
UL

13. ABSTRACT (Maximum 200 words)

Time series modeling as the sum of an autoregressive (AR) process and sinusoids is proposed. When the AR model order is infinite, it is called Canonical Autoregressive Decomposition (CARD) and is equivalent to the Wold decomposition. Maximum likelihood estimation of the sinusoidal and AR parameters is shown to require minimization with respect to only the unknown frequencies. Although the estimation problem is nonlinear in the sinusoidal amplitudes and AR parameters, it was reduced to a linear least squares problem by using a nonlinear paarameter transformation. A general class of signals for which such parameter transformations are applicable, thereby reducing estimator complexity drastically, is derived. This class includes sinusoids as well as polynomials and polynomial-times-exponential signals. The ideas are based on the theory of invariant subspaces for linear operators. CARD serves as a powerful modeling tool in signal plus noise settings and therefore finds application in a large variety of statistical signal processing problems.

14. SUBJECT TERMS

93-23113

15. NUMBER OF PAGES
42

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR(SAME AS REPORT)

NSN 7540-01-280-5500

93 10 1 230

Standard Form 298 (Rev. 2-89)

UNIVERSITY OF RHODE ISLAND

AFOSR-91-0024
1 OCT 90 - 30 SEP 92

SPECTRAL ANALYSIS ON THE CANONICAL AUTOREGRESSIVE DECOMPOSITION

Steven Kay† and Venkatesh Nagesha†
Dept. of Electrical Engineering
University of Rhode Island
Kingston, RI 02881

Permission to publish abstract separately is granted

Abstract

Time series modeling as the sum of an autoregressive (AR) process and sinusoids is proposed. When the AR model order is infinite, we call this the Canonical Autoregressive Decomposition (CARD) and is equivalent to the Wold decomposition. Maximum likelihood estimation of the sinusoidal and AR parameters is shown to require minimization with respect to only the unknown frequencies. Although the estimation problem is nonlinear in the sinusoidal amplitudes and AR parameters, we reduce it to a linear least squares problem by using a nonlinear parameter transformation. A general class of signals for which such parameter transformations are applicable, thereby reducing estimator complexity drastically, is derived. This class includes sinusoids as well as polynomials and polynomial–times–exponential signals. The ideas are based on the theory of invariant subspaces for linear operators. CARD serves as a powerful modeling tool in signal plus noise settings and therefore finds application in a large variety of statistical signal processing problems. We briefly discuss some applications such as spectral analysis, broadband/transient detection using line array data, and fundamental frequency estimation for periodic signals.

*EDICS Category 5.1 or 5.2.1

†This work was supported by the Naval Underwater Systems Center, Newport, RI under contract N66604-91-M-BW47 and by the Air Force Office of Scientific Research under contract AFOSR-91-0024

1 Introduction

Many problems encountered in statistical signal processing may be posed as one that attempts to decompose a time series into its signal and noise components. A common example is spectral analysis in which we interested in extracting sinusoidal components from a background of noise. More generally, this is the problem of estimation of a mixed spectrum, one that is composed of continuous as well as discrete components. Many approaches to this problem have been tried but the results have been generally unsatisfactory. The time series models appear to be either sinusoids in *white* noise, used in eigenanalysis approaches, or *colored noise only*, used in autoregressive / autoregressive moving average modeling for spectral estimation.

The problem of statistical inference for a mixed spectrum is not new. Early attempts by Whittle can be found in [23] and a more general description in [16]. The methods assume that the sinusoids are well resolved and that the sinusoidal parameter estimates are not influenced by the coloration of the background noise. This facilitates the implementation of two step procedures in which the sinusoids are first estimated and removed, followed by spectral estimation of the background noise. Such an approach is justified by appealing to asymptotic (as the data record length increases) arguments. While the assumption of sinusoidal resolution is easily understood in asymptotic arguments, the second assumption, i.e., that the possible coloration of the noise can be neglected (or, the noise can be assumed to be white) in estimating the sinusoidal parameters, stems from a result in [9]. The reasoning is that the asymptotic performance of sinusoidal parameter estimators based on colored or white noise assumptions are identical. However, in most practical problems of interest herein, these asymptotic assumptions are invalid due to the availability of only very short data records. The more important problem from a practical viewpoint, is to be able to *jointly* estimate the sinusoidal components as well as the background noise. We propose to do this by employing what we term the canonical autoregressive decomposition (CARD) model.

It should also be noted that the same type of problem arises in regression analysis [20]. There, if the errors are not white, a weighted least squares estimator is preferred. This, however, assumes knowledge of the error covariance structure. When the latter is unknown, the usual practice is to iteratively estimate the error covariance and the regression coefficients, or to ignore the error correlation entirely and implement an ordinary least squares estimator [3, 15, 20]. Neither approach is optimal (for the given problem) in any sense. Again we would

DTIC QUALITY INSPECTED 3

Distribution/
Availability Codes
Avail and/or
Dist Special
A-1

like to be able to jointly estimate the regression parameters (which we may think of as the signal) and the error (noise) covariance structure. The CARD model can be extended to encompass signals (or trends) such as exponential, polynomial, sinusoidal, or combinations of these. These results can be used to solve some regression problems with autocorrelated errors, which are of interest in the social and behavioral sciences [20, Chap. 6].

Finally, the key to the usefulness of our results is that the resulting estimation problem can be reduced to a simple linear least squares problem by a suitable transformation of the parameter space. This drastically reduces the overall complexity of the estimation problem.

The paper begins with an introduction to canonical representations for wide-sense stationary (WSS) processes based on the Wold decomposition. For most practical purposes, any WSS random process can be represented as an infinite-order autoregressive process plus a sum of sinusoids. We term this the canonical autoregressive decomposition. Maximum likelihood estimation of the unknown parameters in the CARD model is the subject of the next section. While this seems to be a highly nonlinear estimation problem, we reduce its complexity by applying a nonlinear parameter transformation. Related results may also be found in [1, 2, 14, 21, 22]. The problem is finally reduced to maximization with respect to only the unknown sinusoidal frequencies. The following section contains generalizations of these ideas to handle deterministic signals (other than sinusoids) in autoregressive noise. These signals include complex exponentials, polynomials, polynomials-times-exponentials. We prove a theorem which yields all possible signal types for which our parameter transformations simplify maximum likelihood estimation. This generalizes the known results for non-zero mean [1, p. 200], pure sinusoids [2, 14, 22], complex exponentials [21], polynomials [4]. In the fifth section, we briefly touch upon some signal processing applications such as broadband/transient detection using line array data, fundamental frequency estimation, spectral analysis, etc. Numerical examples to illustrate some applications are also presented.

2 Canonical Autoregressive Decomposition

The canonical autoregressive decomposition (CARD) model is based on the Wold decomposition, the Lebesgue decomposition, and the Kolmogorov linear prediction theory. We give only the essential theorems to motivate our use towards the CARD model. The interested reader may consult [5, 13, 16, 19] for further details.

By the Wold decomposition theorem, any wide-sense stationary random process can be expressed as,

$$x[n] = s[n] + w[n] \ , \tag{1}$$

where

1. $s[n]$ and $w[n]$ are uncorrelated processes
2. $s[n]$ is a singular process in that it can be perfectly predicted by a linear combination of its past values.
3. $w[n]$ is a regular process (cannot be perfectly predicted by a linear combination of its past samples) having a moving-average representation

$$w[n] = \sum_{k=0}^{\infty} b[k]u[n-k] \tag{2}$$

with $\sum_{k=0}^{\infty} \mid b[k] \mid^2 < \infty$, and $u[n]$ being a white noise process uncorrelated with $s[n]$.

By the Lebesgue decomposition theorem, the spectral distribution function (which may be thought of as the integrated power spectral density) of any wide-sense stationary random process can be decomposed as

$$S(f) = S_1(f) + S_2(f) + S_3(f) \ , \tag{3}$$

where $S_1(f)$ is an absolutely continuous distribution function, $S_2(f)$ is a step function with steps P_i at frequencies f_i, and $S_3(f)$ is a singular function. For all practical purposes, the third component $S_3(f)$ can be ignored [13, 16]. By the Wiener-Khintchine theorem, the autocorrelation function corresponding to $S_2(f)$ is

$$\begin{aligned} r_2[m] &= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{j2\pi fm} dS_2(f) \\ &= \sum_{i=1}^{s} P_i e^{j2\pi f_i m} \ , \end{aligned}$$

or $s_2[n]$ is a sum of sinusoids and is perfectly predictable. Finally, the absolutely continuous component could represent either a regular, or singular process, depending on whether [16], [23, App. 2]

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln S_1'(f) df > -\infty \tag{4}$$

holds or not, respectively. Here $'$ denotes differentiation, so that $S_1'(f)$ denotes the PSD corresponding to the absolutely continous component in (3). For the most part, the singular process therein corresponds to perfect prediction based on strictly-infinite past. Examples of such processes include perfectly band-limited processes. By assuming that the absolutely continuous component represents a regular process, i.e., (4) is satisfied (which identifies $S_1(f)$ in (3) with $w[n]$ in (1)), we rewrite (1) as

$$x[n] = \sum_{k=0}^{\infty} b[k]u[n-k] + \sum_{k=1}^{s} A_i e^{j2\pi f_i n} \tag{5}$$

where A_i's are zero mean, complex-valued, random variables uncorrelated with each other and with the $u[n]$'s, and $\mathcal{E}(|A_i|^2) = P_i$.

The CARD model is finally obtained by noting that, under some conditions, an infinite-order moving average process is equivalent to an infinite-order autoregressive process. That this is not true in general is illustrated by [13] the moving average (MA) process

$$w[n] = u[n] - u[n-1] \ ,$$

which does not have a autoregressive (AR) representation. The reason stems from the fact that the MA process has a zero on the unit circle, so that the transfer function

$$\frac{1}{B(z)} = \frac{1}{1 - z^{-1}} \ ,$$

is not analytic for $z \geq 1$. To avoid this problem, we assume that the power spectral density (PSD) of the MA component in (5) is bounded away from zero, or that $P_x(f) \geq \zeta > 0$ for all f. Then, (4) is satisfied. This assumption is not overly restrictive in that all physical processes are subject to observation noise, causing the PSD to be strictly positive. Note that the general conditions for existence of infinite-order AR representations are extremely complicated [13].

In summary, we have the following results

Theorem 1 *If $x[n]$ is any wide-sense stationary process with PSD $P_x(f) \geq \zeta > 0$ for all f so that equation (4) is satisfied, then it can be decomposed as*

$$x[n] = w[n] + \sum_{i=1}^{s} A_i e^{j2\pi f_i n} \ , \tag{6}$$

where $w[n]$ is an infinite-order AR process given by

$$w[n] = -\sum_{k=1}^{\infty} a[k]w[n-k] + u[n] \ . \tag{7}$$

The $u[n]$ process is white noise with variance σ^2, and uncorrelated with the A_i's. The A_i's are zero mean, complex-valued, uncorrelated random variables with variances, $\mathcal{E}(|A_i|^2) = P_i$.

As a result of this decomposition, termed the canonical autoregressive decomposition, we have the PSD

$$P_x(f) = \frac{\sigma^2}{|\mathcal{A}(e^{j2\pi f})|^2} + \sum_{i=1}^{s} P_i \delta(f - f_i) \ , \tag{8}$$

where $\mathcal{A}(e^{j2\pi f}) = \sum_{k=0}^{\infty} a[k]e^{-j2\pi fk}$ and $a[0] = 1$. The word "canonical" stems from the fact that such a decomposition can represent the second-order moment properties of any physical process encountered in practice. Lastly, in most applications, we will assume that the order of the AR process is finite so that $\mathcal{A}(e^{j2\pi f})$ in (8) will be $\mathcal{A}(e^{j2\pi f}) = 1 + \sum_{k=1}^{p} a[k]e^{-j2\pi fk}$. Hence, the unknown parameters are $\{p, a[1], a[2], \ldots, a[p], \sigma^2, s, P_1, f_1, \ldots, P_s, f_s\}$.

3 Parameter Estimation

We will now derive an approximate maximum likelihood estimator (MLE) of the CARD model parameters. It is assumed that the AR process is Gaussian but that the sinusoidal amplitudes are *deterministic*. Instead of attempting to estimate the powers P_i in the CARD model, we modify the problem by assuming that the A_i's are unknown, deterministic constants. The reason for this is that the MLE for the original problem is highly intractable. Furthermore, since only a single realization of the time series is available, it makes no sense to estimate the variances P_i, which are ensemble averages (the MLE will be inconsistent). Finally, we will also assume that the number of sinusoids (s) and the AR model order (p) are known. Therefore, the unknown parameters are $\{a[1], a[2], \ldots, a[p], \sigma^2, A_1, f_1, \ldots, A_s, f_s\}$.

To summarize the estimation problem, we assume the time series model to be

$$x[n] = s[n] + w[n] \ , \tag{9}$$

where

$$s[n] = \sum_{i=1}^{s} A_i e^{j2\pi f_i n} \ , \tag{10}$$

$$w[n] = -\sum_{k=1}^{p} a[k]\, w[n-k] + u[n] \ . \tag{11}$$

Here $u[n]$ is complex, white Gaussian with variance σ^2 so that $w[n]$, and hence $x[n]$, is complex Gaussian. The unknown parameters are $\boldsymbol{\theta} = \left[\mathbf{A}^T\, \mathbf{f}^T\, \mathbf{a}^T\, \sigma^2\right]^T$, where $\mathbf{A} = [A_1\, A_2 \ldots A_s]^T$, $\mathbf{f} = [f_1\, f_2 \ldots f_s]^T$, and $\mathbf{a} = \left[a[1]\, a[2] \ldots a[p]\right]^T$.

Assuming that a finite-length data record $\mathbf{x}_o = \left[x[0]x[1]\ldots x[N-1]\right]^T$ is available, the approximate (actually conditional) PDF [8, 10] is given by

$$p(\mathbf{x}_o;\boldsymbol{\theta}) = \frac{1}{(\pi\sigma^2)^{N'}} \exp\left\{-\frac{1}{\sigma^2}\sum_{n=p}^{N-1}\left|\sum_{k=0}^{p} a[k]\,(x[n-k] - s[n-k])\right|^2\right\} \ , \tag{12}$$

where $N' = N - p$ and $a[0] = 1$. Maximizing the PDF with respect to σ^2 leads to

$$\hat{\sigma}^2 = \frac{1}{N'} J(\boldsymbol{\theta}') = \frac{1}{N'}\sum_{n=p}^{N-1}\left|\sum_{k=0}^{p} a[k]\Big(x[n-k] - s[n-k]\Big)\right|^2 \tag{13}$$

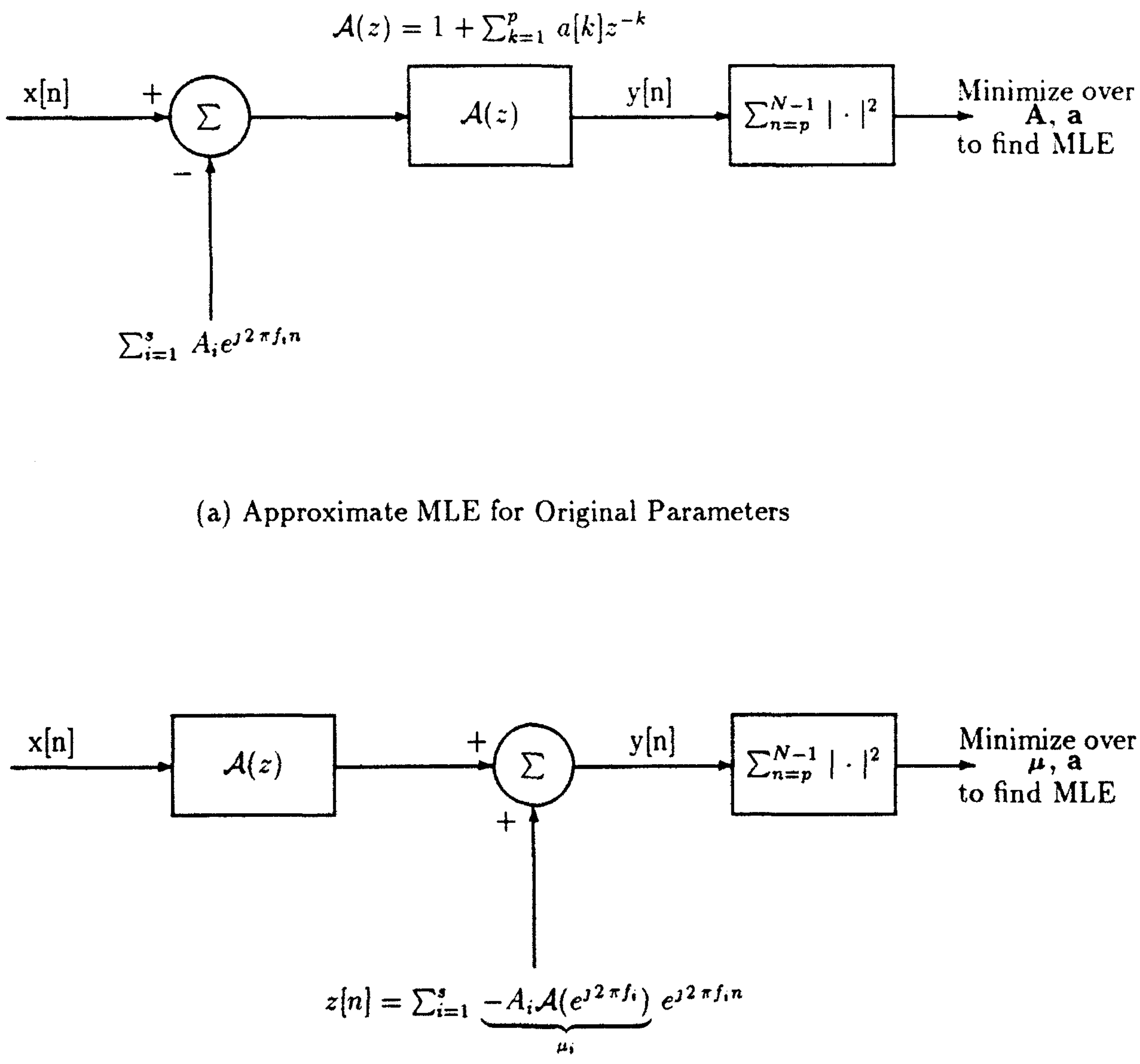
so that to find the MLE of $\boldsymbol{\theta}'$ ($\boldsymbol{\theta}$ without σ^2), we must now minimize $J(\cdot)$. Using (10), we rewrite J as

$$\begin{aligned}
J(\boldsymbol{\theta}') &= \sum_{n=p}^{N-1}\left|\sum_{k=0}^{p} a[k]\left(x[n-k] - \sum_{i=1}^{s} A_i e^{j2\pi f_i(n-k)}\right)\right|^2 & \text{(14a)}\\
&= \sum_{n=p}^{N-1}\left|\sum_{k=0}^{p} a[k]x[n-k] - \sum_{i=1}^{s} A_i \sum_{k=0}^{p} a[k]e^{j2\pi f_i(n-k)}\right|^2 & \\
&= \sum_{n=p}^{N-1}\left|\sum_{k=0}^{p} a[k]x[n-k] - \sum_{i=1}^{s} A_i \left(\sum_{k=0}^{p} a[k]e^{-j2\pi f_i k}\right) e^{j2\pi f_i n}\right|^2 & \\
&= \sum_{n=p}^{N-1}\left|\sum_{k=0}^{p} a[k]x[n-k] + \sum_{i=1}^{s} \mu_i\, e^{j2\pi f_i n}\right|^2 \ , & \text{(14b)}
\end{aligned}$$

where $\mu_i = -A_i \sum_{k=0}^{p} a[k]\, e^{-j2\pi f_i k}$. Clearly J in (14b) is quadratic with respect to the $a[k]$'s and μ_i's. Thus a nonlinear (due to interaction between $a[k]$ and A_i in (14a)) least squares problem is reduced to a linear least squares problem by the nonlinear parameter transformation,

$$\mu_i = -A_i\left(\sum_{k=0}^{p} a[k]\, e^{-j2\pi f_i k}\right) \ . \tag{15a}$$

The equivalence of the problems in (14a) and (14b) via the parameter transformation is illustrated in Figure 1.



(a) Approximate MLE for Original Parameters

(b) Approximate MLE for Transformed Parameters

Figure 1: Approximate MLE and Sinusoidal Amplitude Transformation

Note that $-\mu_i$ can be thought of as the amplitude of the prewhitenened sinusoidal signal, the prewhitener being implementable if the AR parameters were known. The transformation from A_i to μ_i is one-to-one iff the AR process does not have poles on the unit circle (which of course it does not), in which case we obtain

$$A_i = -\mu_i \Big/ \left(\sum_{k=0}^{p} a[k]\, e^{-j 2\pi f_i k} \right) . \tag{15b}$$

Now, define

$$\begin{aligned}
\mathbf{x} &= \left[x[p]\, x[p+1] \cdots x[N-1] \right]^T \\
\mathbf{H} &= \begin{bmatrix} x[p-1] & x[p-2] & \cdots & x[0] \\ x[p] & x[p-1] & \cdots & x[1] \\ \vdots & \vdots & \ddots & \vdots \\ x[N-2] & x[N-3] & \cdots & x[N-1-p] \end{bmatrix} \quad (N-p) \times p \\
\mathbf{E} &= \begin{bmatrix} e^{j2\pi f_1 p} & e^{j2\pi f_2 p} & \dots & e^{j2\pi f_s p} \\ e^{j2\pi f_1 (p+1)} & e^{j2\pi f_2 (p+1)} & \dots & e^{j2\pi f_s (p+1)} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j2\pi f_1 (N-1)} & e^{j2\pi f_2 (N-1)} & \dots & e^{j2\pi f_s (N-1)} \end{bmatrix} \quad (N-p) \times s
\end{aligned} \tag{16}$$

so that (14b) becomes

$$J'(\mathbf{a}, \boldsymbol{\mu}, \mathbf{f}) = \|\mathbf{x} + \mathbf{H}\mathbf{a} + \mathbf{E}\boldsymbol{\mu}\|^2 , \tag{17}$$

where $\boldsymbol{\mu} = [\mu_1\, \mu_2 \, \dots \, \mu_s]^T$, and $\|\cdot\|$ denotes the Euclidean norm of a complex vector. It can be shown that the concatenated matrix $[\mathbf{H}\,\mathbf{E}]$ is full rank with probability one if the data conforms to the assumed model and $N \geq 2p + s$, hence a unique minimum exists for $J'(\cdot)$. Minimizing J' with respect to $\mathbf{a}$ and $\boldsymbol{\mu}$ leads to (see Appendix A for details)

$$\begin{aligned}
J''(\mathbf{f}) &= \mathbf{x}^H \left[\mathbf{P}_E^{\perp} - \mathbf{P}_E^{\perp}\mathbf{H} \left(\mathbf{H}^H \mathbf{P}_E^{\perp} \mathbf{H} \right)^{-1} \mathbf{H}^H \mathbf{P}_E^{\perp} \right] \mathbf{x} & \text{(18a)} \\
&= \mathbf{x}^H \left[\mathbf{P}_H^{\perp} - \mathbf{P}_H^{\perp}\mathbf{E} \left(\mathbf{E}^H \mathbf{P}_H^{\perp} \mathbf{E} \right)^{-1} \mathbf{E}^H \mathbf{P}_H^{\perp} \right] \mathbf{x} & \text{(18b)}
\end{aligned}$$

where $\mathbf{P}_E^{\perp} = \mathbf{I} - \mathbf{E}\left(\mathbf{E}^H\mathbf{E}\right)^{-1}\mathbf{E}^H$ and $\mathbf{P}_H^{\perp} = \mathbf{I} - \mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H$ are projection matrices. All the required inverses will exist since $[\mathbf{H}\,\mathbf{E}]$ is full rank with probability one. The corresponding $\hat{\mathbf{a}}$ and $\hat{\boldsymbol{\mu}}$ are given by either

$$\begin{aligned}
\hat{\mathbf{a}} &= -\left(\mathbf{H}^H \mathbf{P}_E^{\perp} \mathbf{H} \right)^{-1} \mathbf{H}^H \mathbf{P}_E^{\perp} \mathbf{x} & \text{(19a)} \\
\hat{\boldsymbol{\mu}} &= -\left(\mathbf{E}^H \mathbf{E} \right)^{-1} \mathbf{E}^H \left(\mathbf{x} + \mathbf{H}\hat{\mathbf{a}} \right) , & \text{(19b)}
\end{aligned}$$

or

$$\hat{\mu} = -\left(\mathbf{E}^H\mathbf{P}_H^\perp\mathbf{E}\right)^{-1}\mathbf{E}^H\mathbf{P}_H^\perp\mathbf{x} \tag{20a}$$

$$\hat{\mathbf{a}} = -\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\left(\mathbf{x}+\mathbf{E}\hat{\mu}\right) . \tag{20b}$$

The estimation of the unknown frequencies is accomplished by minimizing $J''(\cdot)$ in (18) and will in general require iterative, nonlinear optimization techniques. Using these estimated frequencies in (19) or (20), and (15b), (13), the other parameters can be obtained. Equations (18) - (20) have some interesting interpretations. The estimate $\hat{\mathbf{a}}$ in (19a) can be thought of as the usual covariance method *after subtracting* the signal components (via $\mathbf{P}_E^\perp$) in the data vector $\mathbf{x}$ and the data matrix $\mathbf{H}$. The resulting minimum value for the prediction error is given in (18a), which must be minimized to obtain the unknown frequencies. Finally, $\hat{\mu}$ in (19b) is the signal amplitude estimate based on the 'prediction error' $\mathbf{x}+\mathbf{H}\hat{\mathbf{a}}$. Similar interpretations are possible for the other expressions wherein the $\mathbf{P}_H^\perp$ operator is interpreted as an AR prewhitener, but one that is based on signal + noise.

We now give an example. Assume that we wish to estimate the frequency of a complex sinusoid in AR noise. Then, using $s = 1$ in (18b), we get (since $\mathbf{E}^H\mathbf{P}_H^\perp\mathbf{E}$ is a scalar)

$$J''(f) = \mathbf{x}^H\mathbf{P}_H^\perp\mathbf{x} - \frac{|\,\mathbf{e}^H\mathbf{P}_H^\perp\mathbf{x}\,|^2}{\mathbf{e}^H\mathbf{P}_H^\perp\mathbf{e}}$$

where $\mathbf{e} = \left[e^{j2\pi fp}\, e^{j2\pi f(p+1)} \,\ldots\, e^{j2\pi f(N-1)}\right]^T$, $\mathbf{P}_H^\perp = \mathbf{I} - \mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H$ and $\mathbf{x}$, $\mathbf{H}$ are given in (16). Hence an approximate MLE of the frequency must maximize

$$\zeta(f) = \frac{|\,\mathbf{e}^H\mathbf{P}_H^\perp\mathbf{x}\,|^2}{\mathbf{e}^H\mathbf{P}_H^\perp\mathbf{e}} . \tag{21}$$

This has some interesting interpretations. Let $\hat{\mathbf{u}}_0 = \mathbf{P}_H^\perp\mathbf{x} = \mathbf{x}+\mathbf{H}\hat{\mathbf{a}}_0$, where $\hat{\mathbf{a}}_0$ is the usual least squares estimator (covariance method) of the AR parameters, and $\hat{\mathbf{u}} = \mathbf{x}+\mathbf{H}\hat{\mathbf{a}}$, where $\hat{\mathbf{a}}$ is the estimate of the AR parameters given by (20b). Clearly, both are estimates of the driving noise sequence. As shown in Appendix B, we can rewrite (21) as,

$$\zeta(f) = \frac{\left(\mathbf{e}^H\hat{\mathbf{u}}_0\right)^*\mathbf{e}^H\hat{\mathbf{u}}}{\mathbf{e}^H\mathbf{e}} . \tag{22}$$

Noting that for low SNR, $\hat{\mathbf{a}}_0 \approx \hat{\mathbf{a}}$, so that $\hat{\mathbf{u}}_0 \approx \hat{\mathbf{u}}$, we obtain

$$\begin{aligned}\zeta(f) &\approx \frac{1}{N-p}\,|\,\mathbf{e}^H\hat{\mathbf{u}}_0\,|^2 \\ &= \frac{1}{N-p}\,|\,U_0(f)\,|^2\end{aligned}$$

where $U_0(f) = \mathbf{e}^H \hat{\mathbf{u}}_0$ denotes the Fourier transform of the driving noise sequence estimate. Neglecting end effects (for $N \gg p$), we get $U_0(f) = \hat{A}_0(e^{j2\pi f})X(f)$, where $\hat{A}_0(e^{j2\pi f}) = 1 + \sum_{k=1}^{p} \hat{a}_0[k]e^{-j2\pi f k}$, and $X(f) = \sum_{n=0}^{N-1} x[n]e^{-j2\pi f n}$ denotes the Fourier transform of $x[n]$. Hence we obtain,

$$\zeta(f) \;\approx\; \frac{1}{N} \mid X(f)\hat{A}_0(e^{j2\pi f}) \mid^2 ,$$

or equivalently, we may maximize

$$\zeta'(f) \;=\; \frac{\mid X(f) \mid^2 /N}{\hat{\sigma}_0^2 / \mid \hat{A}_0(e^{j2\pi f}) \mid^2} ,$$

where $\hat{\sigma}_0^2$ is the MLE of the driving noise variance assuming no signal is present. But, the numerator is just the periodogram $I_x(f)$ and the denominator, the estimated AR PSD $\hat{P}_{AR}(f)$. Hence for low SNR, the approximate MLE maximizes

$$\zeta'(f) \;=\; \frac{I_x(f)}{\hat{P}_{AR}(f)} .$$

We may view the procedure as either prewhitening the data followed by peak picking the periodogram, or normalizing the periodogram by the background noise followed by peak picking.

4 Extensions to Other Signals

The key to analytical solutions in the previous section was the parameter transformation (15a) which rendered the problem linear in the new parameters. A careful consideration of the transformation reveals that this was possible because sinusoids are eigenfunctions of a linear time-invariant (LTI) FIR filter. If the input to an LTI FIR filter is a causal sinusoidal sequence, then the output (after disregarding the initial samples) is also sinusoidal, but with a different amplitude. More generally, if the input sequence belongs to a subspace spanned by sinusoidal sequences (as in (10)), the output sequence also lies in the same subspace. This property is an instance of the so-called subspace invariance of linear transformations [6, 7]. However, subspaces spanned by eigenfunctions are only the most primitive of invariant subspaces; more generally they involve generalized eigenfunctions and Jordan chains [7].

We now state our results which extend the case of sinusoids in AR noise to more general signals in AR noise. Some of these signals are

1. damped and undamped sinusoids: $s[n] = \sum_{i=1}^{s} A_i\, r_i^n\, e^{j2\pi f_i n}$

2. polynomials: $s[n] = \sum_{i=1}^{s} A_i\, n^{i-1}$

3. polynomials-times-exponentials: $s[n] = r^n\, e^{j2\pi f n}\,(\sum_{i=1}^{s} A_i\, n^{i-1})$,

or their linear combinations.

To do so, we present some definitions followed by our theorems (the proofs of which are in Appendices C and D). The definitions and theorems are designed to answer the following question:

For $s[n]$ given by

$$s[n] = \sum_{i=1}^{s} A_i s_i[n] \ ,$$

where $s_i[n]$ are known and the amplitudes are unknown, we wish to replace the minimization of

$$\begin{aligned} J(\theta') &= \sum_{n=p}^{N-1} \left| \sum_{k=0}^{p} a[k]\,(x[n-k] - s[n-k]) \right|^2 \\ &= \sum_{n=p}^{N-1} \left| \sum_{k=0}^{p} a[k]x[n-k] \ - \ \sum_{k=0}^{p} a[k] \sum_{i=1}^{s} A_i s_i[n-k] \right|^2 \end{aligned}$$

by the minimization of

$$J(\theta') = \sum_{n=p}^{N-1} \left| \sum_{k=0}^{p} a[k]x[n-k] \ + \ \sum_{i=1}^{s} \mu_i s_i[n] \right|^2 \ .$$

Hence, we need to have that for all $\mathbf{A}$, $\mathbf{a}$,

$$\sum_{k=0}^{p} a[k] \sum_{i=1}^{s} A_i s_i[n-k] = -\sum_{i=1}^{s} \mu_i s_i[n] \ , \tag{23}$$

for $n = p, p+1, \ldots, (N-1)$. To be useful, we require the transformation in (23) to hold for any N and any p for which the maximum likelihood problem is well defined. In effect, we require that any linear combination of the basis signals $\{s_1[n], s_2[n], \ldots, s_s[n]\}$ when transformed by the linear time-invariant filter $\mathcal{A}(z) = 1 + \sum_{k=1}^{p} a[k]z^{-k}$ must produce a signal that is again a linear combination of the basis signals, $\{s_1[n], s_2[n], \ldots, s_s[n]\}$. Symbolically, letting L denote the linear filter operation by $\mathcal{A}(z)$, we require that

$$L\left\{\sum_{i=1}^{s} A_i s_i[n]\right\} = \sum_{i=1}^{s} D_i s_i[n] \quad \text{for } p \le n \le N-1 \ . \tag{24}$$

We now find all possible basis signal sets for which (24) holds. Note that it must hold for all L, or equivalently for all $\mathbf{a}$. The B_i's will of course depend on the values of $\mathbf{a}$ and $\mathbf{A}$.

Definition 1 *C_N denotes the linear vector space of all length-N sequences. An element in the space will be denoted by $z[0]$, $z[1]$, $\ldots$, $z[N-1]$.*

Definition 2 *Let $\mathcal{L}$ denote the set of all p-th order linear transformations on C_N. An element in $\mathcal{L}$, described by the parameters $a[1]$, $a[2]$, $\ldots$, $a[p]$, with $a[p] \neq 0$, operates on an element $z[0]$, $z[1]$, $\ldots$, $z[N-1]$ in C_N to produce an element $y[p]$, $y[p+1]$, $\ldots$, $y[N-1]$ in C_{N-p} according to*

$$y[n] = z[n] + \sum_{k=1}^{p} a[k]z[n-k] \quad , for \quad p \leq n \leq N-1 \tag{25}$$

Definition 3 *A p-shift operator $\mathcal{P}(\cdot)$ operates on an element $z[0]$, $z[1]$, $\ldots$, $z[N-1]$ in C_N to produce an element $z[p]$, $z[p+1]$, $\ldots$, $z[N-1]$ in C_{N-p}. $\mathcal{P}$ operates similarly on subspaces in C_N.*

Definition 4 *A subspace $\mathcal{S}$ in C_N is said to be invariant with respect to $\mathcal{L}$ if $L : \mathcal{S} \rightarrow \mathcal{P}(\mathcal{S})$ for any $L \in \mathcal{L}$.*

With these definitions, we have our main results. Some related results for $p = 1$ are alluded to in [7, p. 200].

Theorem 2 *The invariant subspace $\mathcal{S}$ of dimension $M \leq N - 2p + 1$ is composed of r invariant subspaces $\mathcal{S}_1$, $\mathcal{S}_2$, $\ldots$, $\mathcal{S}_r$ where $\mathcal{S} = \mathcal{S}_1 \oplus \mathcal{S}_2 \oplus \cdots \oplus \mathcal{S}_r$, and $\oplus$ denotes the direct sum of the subspaces. The basis for $\mathcal{S}_i$, which is of dimension m_i, is*

$$\lambda_i^n, \binom{n}{1} \lambda_i^{n-1}, \ldots, \binom{n}{m_i - 1} \lambda_i^{n-(m_i-1)} \tag{26}$$

for $n = 0, 1, \ldots, (N-1)$, and where λ_i is any nonzero complex number and $\sum_{i=1}^{r} m_i = M$.

Proof: See Appendix C. ◇

Note that for the problem of (17) to be well defined, we must have $s \leq N - 2p$. This is more restrictive than Theorem 2 (where $M \leq N - 2p + 1$), so that the parameter transformation itself imposes no new problem constraint.

Theorem 3 *Let $s[n]$ belong to the invariant subspace $\mathcal{S}$. From Theorem 2, the signal is of the form*

$$s[n] = \sum_{i=1}^{r} \sum_{j=0}^{m_i-1} B_j^{(i)} \binom{n}{j} \lambda_i^{n-j} \ , \tag{27}$$

where $B_j^{(i)}$ are arbitrary constants. Then, for all p-th order AR processes with poles not located at $\lambda_1, \lambda_2, \ldots, \lambda_r$, we have that, for $p \leq n \leq N-1$,

$$\sum_{k=0}^{p} a[k]s[n-k] = \sum_{i=1}^{r} \sum_{j=0}^{m_i-1} D_j^{(i)} \binom{n}{j} \lambda_i^{n-j} \ . \tag{28}$$

The $D_j^{(i)}$ are found using, $\mathbf{D}_i = \mathbf{T}_i\mathbf{B}_i$, for $i = 1, 2, \ldots, r$, where

$$\mathbf{D}_i = \left[D_0^{(i)} \ D_1^{(i)} \ \cdots \ D_{m_i-1}^{(i)}\right]^T \ ,$$

$$\mathbf{T}_i = \sum_{k=0}^{p} a[k]\mathbf{J}_i^{-k} \ , \tag{29}$$

$$\mathbf{J}_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix} \ m_i \times m_i$$

$$\mathbf{B}_i = \left[B_0^{(i)} \ B_1^{(i)} \ \cdots \ B_{m_i-1}^{(i)}\right]^T \ .$$

Proof: See Appendix *D*. ◇

Note that the transformation from $\mathbf{B}_i$ to $\mathbf{D}_i$ is invertible since $\mathbf{T}_i$ is nonsingular. This is because $\mathbf{J}_i^{-k}$ is upper-triangular with λ_i^{-k} along the main diagonal, so that $\mathbf{T}_i$ is also upper-triangular with $\sum_{k=0}^{p} a[k]\lambda_i^{-k}$ along its main diagonal. Since the AR process does not have poles at the λ_i's, $\mathbf{T}_i$ will be nonsingular. This means that the least squares minimization can be done with respect to $\mathbf{D}_i$'s.

We now present some examples and then discuss the restrictions on the modes, λ_i's.

<u>Examples:</u>

1. Sinusoids and damped sinusoids:
   Let $m_i = 1$ and $\lambda_i = r_i e^{j2\pi f_i}$ in (27). Then, we get

$$s[n] = \sum_{i=1}^{r} B^{(i)} r_i^n e^{j2\pi f_i n}$$

2. Polynomials:
   Let $r = 1$, $m_1 = M$, and $\lambda_1 = 1$ in (27). Then, we get

$$s[n] = \sum_{j=0}^{M-1} B_j \binom{n}{j} .$$

   But, $\binom{n}{j} = n!/j!(n-j)! = n(n-1)\cdots(n-j+1)/j!$, so that $\binom{n}{j}$ is seen to be a polynomial of degree j. It can be shown that

$$s[n] = \sum_{j=0}^{M-1} B'_j n^j .$$

   This case was originally discovered by Djuric [4].

3. Polynomials–times–Exponentials:
   Let $r = 1$, $m_1 = M$ and $\lambda_1 = re^{j2\pi f}$ in (27). Then, we get

$$s[n] = \sum_{j=0}^{M-1} B_j \binom{n}{j} (re^{j2\pi f})^{n-j} ,$$

   or

$$s[n] = r^n e^{j2\pi f n} \sum_{j=0}^{M-1} B''_j n^j .$$

Also, any linear combination of these signals is possible.

Once the signals have been chosen, the general form of the output signal is known except for the amplitudes, which are to be estimated. To recover the input signal amplitudes, we use from Theorem 3,

$$\mathbf{B}_i = \mathbf{T}_i^{-1}\mathbf{D}_i .$$

Since $\mathbf{T}_i$ is upper-triangular with $\sum_{k=0}^{p} a[k]\lambda_i^{-k}$ along the diagonal, $\mathbf{T}_i$ will be non-singular iff

$$\mathcal{A}(\lambda_i) = 1 + \sum_{k=1}^{p} a[k]\lambda_i^{-k} \neq 0 , \tag{30}$$

Therefore, for the transformation in (29) to be invertible, the zeros of the filter $\mathcal{A}(z)$, i.e., the poles of the AR process, should not lie at the signal modes. Otherwise, $\mathcal{A}(z)$ annihilates the corresponding signal mode and the original amplitude is not recoverable. In section 3, we had $s[n] = \sum_{i=1}^{s} A_i e^{j2\pi f_i n}$, so that (30) meant that $\mathcal{A}(e^{j2\pi f_i}) \neq 0$, or the poles of the noise process should not coincide with the sinusoidal frequency locations.

5 Some Applications

The CARD model is useful in that it reduces a nonlinear minimization problem involving the AR parameters, signal amplitudes, and signal modes, to one involving only the signal modes. Additionally, because of its form, the signal and noise are independently parameterized. the Fisher information matrix is block diagonal [24]. This has important implications because it says that the Cramer-Rao (CR) bound for the signal parameters is the same for the cases of known AR parameters and unknown AR parameters. For instance, in the spectral analysis problem of section 3, the CR bound for the sinusoidal parameters can be found in [18], although it was not as thoroughly investigated as the white noise case. When the CR bound is attained, the estimation performance is as good as if the AR parameters are known. In detection applications, this translates into optimal detection in unknown AR noise [11].

Although the number of applications of the CARD model and its extensions is quite large, we now briefly describe some of current interest to the signal processing community. They are

1. spectral analysis
2. broadband / transient detection
3. fundamental frequency estimation

A. Spectral Analysis: Spectral analysis of time series using the CARD model can be implemented if the number of sinusoids and the order of the AR process are known. Otherwise, some means of model order selection must be employed. We are currently investigating this problem. Assuming known model orders, the sinusoidal frequencies can be estimated by minimizing $J''(\cdot)$ in (18). Then the AR parameter and sinusoidal frequency estimates are found from (19) or (20). The minimization of J'' will be difficult for more than a few sinusoids, so that we are currently investigating various approaches.

We now present some numerical examples using $N = 25$ data samples consisting of two sinusoids in colored 2^{nd} order AR process. Three generic spectral estimators are considered, one is a purely continuous PSD estimator, the second a purely discrete (line) spectral estimator, and the third is a mixed spectral estimator given by the CARD model. The purpose is to provide some brief numerical comparisons on how the CARD model estimates compare with usual methods when (i) an estimate of the entire spectrum is of interest, and (ii) only the frequency estimates are of interest. The purely continuous PSD estimator chosen here is a 12th order AR spectral estimate computed using the modified covariance method [10]. This is a common choice in such mixed-spectrum problems. The line spectral estimator was obtained by minimizing (18b) with $p = 0$ and $s = 2$, and is a implementation of the MLE for the sinusoids in white noise problem [10]. Strictly speaking, as an estimate of the overall spectrum, this (in itself) will provide a very poor estimate, and higher orders (i.e., $s > 2$) will be needed for this purpose. In applications wherein only the frequencies are of interest, this (along with AR spectral analysis) is the most common implementation. Finally, the CARD estimate is obtained by minimizing (18b) with $p = 2$ and $s = 2$. The AR parameter estimates in CARD are then computed using the estimated frequencies in (19a). The minimization of (18b), in both estimators, was carried out by implementing a 500×500 grid search for the unknown frequencies. The sinusoidal components are indicated by a single line of height A_i^2/σ^2 (the true values of A_i, σ^2 are used, not their estimates) located at the estimated frequency. The display is therefore indicative of the sinusoidal power relative to the noise PSD. Typically, several realizations of these estimators will be plotted so that small variations in the frequency estimates cause the resulting lines to appear adjacent to each other and thereby forming a *narrow band*.

As the first example, a narrowband AR(2) process with poles at $0.95\ e^{j2\pi 0.1}$ and $0.95\ e^{j2\pi 0.2}$ and a driving noise variance of $\sigma^2 = 1.0$ is considered. The sinusoidal frequencies were $f_1 = 0.60$ and $f_2 = 0.62$, and the amplitudes were $A_1 = A_2 = 1.0$. The ***local*** signal to noise ratio (SNR), given by $NA_i^2/P_w(f_i)$, and ***broadband*** SNR, given by $A_i^2/r_w[0]$, are about 25 dB and -17.5 dB, respectively. Figure (2a) shows the CARD spectral estimate for 10 independent realizations of the data set. The line spectrum for the same data set is given in Figure (2b). The frequency estimates are highly biased (as expected, the model is incorrect) and are very close to the peaks in the noise spectrum. This is attributed to the extremely low sinusoidal power. The AR spectral estimate, shown in Figure (2c), exhibits a number of false peaks (as expected, see [10]), and sinusoidal resolution is poor. Finally, the average CARD spectrum (computed using 50 trials) is shown in Figure (2d) and is very close to the true PSD.

To illustrate the effect of sinusoidal frequency location relative to the noise peaks, we change the sinusoidal frequencies to $f_1 = 0.30$ and $f_2 = 0.32$. The local SNR reduces to about 11.5 dB, which is too low for all three algorithms. The local SNR can also be thought of as $N \mid \mu_i \mid^2 / \sigma^2$, where μ_i is the "prewhitened" signal amplitude (see (15a)). Hence by changing the frequencies we are effectively reducing the SNR that dictates performance of the CARD model. The amplitudes were increased to $A_1 = A_2 = 10$, for which the local SNR was about 31.5 dB and the broadband SNR was about 2.5 dB. Figures (3*a*) - (3*c*) show the three spectral estimators for 10 independent realizations. Since the frequency estimates in Figure (3*b*) are a little unclear, Figures (3*d*) and (3*e*) show the frequency estimates using CARD and line spectral estimators for 50 independent realizations. The bias in the line spectral estimate for $\hat{f}_1$ is towards the noise spectral peaks, while that for $\hat{f}_2$ is towards 0.31. No such systematic bias is noticable in the CARD estimates. The average CARD spectrum (computed using 50 trials) is shown in Figure (3*f*) and is quite close to the true PSD.

It is clear that the improvement (vis a vis estimator bias) in the CARD frequency estimate over the usual frequency estimate (i.e., minimizing (18*b*) with $s = 2$ and $p = 0$) is not as drastic as in the previous example. Loosely speaking, the latter estimator searches for total spectral power in narrow bands (sinusoidal power, noise peak power, etc.) Hence, it will yield reasonable (low bias) frequency estimates when the sinusoidal power is much larger or comparable to the noise peaks (as in Example 2), and will yield poor (highly biased) frequency estimates when the noise peaks are much larger than the sinusoidal powers (as in Example 1). An analysis of the CR bound brings out the dependence of the CARD estimator performance (via variance) on the local SNR as well as $A_1^2/P_w'(f)$, where $'$ denotes derivative with respect to f, and is currently in progress.

A third example using broadband AR noise is now considered. It illustrates the effectiveness of the CARD model for narrowband + broadband spectra, which is difficult to implement using other modern spectral analysis methods [10]. The AR(2) poles are located at $0.70\, e^{j2\pi 0.1}$ and $0.70\, e^{j2\pi 0.2}$, and the driving noise variance was $\sigma^2 = 0.5$. The two sinusoidal frequencies were $f_1 = 0.60$ and $f_2 = 0.62$, and the amplitudes were $A_1 = A_2 = 1.0$. The local and broadband SNRs are about 25 dB and 5 dB, respectively. Figure (4*a*) shows the CARD spectral estimate for 10 independent realizations of the data. The line spectral estimate is given in Figure (4*b*) and shows one sinusoid near 0.61 and the other sinusoid in the general vicinity of the noise power concentration (about 0.05 to 0.25). The AR spectral estimates in Figure (4*c*) still show a large number of false peaks, but the sinusoidal resolution is a little better (as compared to Example 1) probably because the broadband SNR is a little higher. Finally, the averaged CARD estimate (computed using 50 trials) is shown in Figure (4*d*) and is quite close to the true PSD.

B. Broadband Signal Detection: The difficulty of optimization with respect to the unknown frequencies can be substantially reduced if the frequencies are known to be related to each other. One case is broadband detection for linear arrays in which a broadband signal can be modeled as a sum of 2-D sinusoids. Since the signal is known to arrive from a certain direction, we know that the frequencies lie along a "bearing line." All that is unknown is the bearing. This results in a relatively simple 1-D optimization over the possible bearings. The interested reader must consult [12] for further details.

C. Fundamental Frequency Estimation: Another related application area is fundamental frequency estimation for a periodic signal in colored noise. This problem arises in speech processing as pitch estimation in colored noise, the co-channel problem (for sum of voiced plus unvoiced speech segments) [17], in biomedical applications as electro-cardiogram monitoring in colored noise, etc. We develop the estimation procedure in some detail.

Assume that we observe a periodic signal in Gaussian AR noise, or

$$x[n] = \sum_{i=1}^{s} A_i e^{j2\pi i f_0 n} + w[n] \ , \tag{31}$$

for $n = 0, 1, \ldots, (N-1)$. The fundamental frequency f_0 is to be estimated along with the Fourier series coefficients A_i and the AR parameters, $a[1]$, $a[2]$, $\ldots$, $a[p]$, σ^2. The number of harmonics s is given by $\lfloor 0.5/f_0 \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. The fact that the number of nuisance parameters $s+p+1$ to be estimated increases with decreasing f_0, can sometimes cause problems with the ML approach. In a sense, the model order changes with a parameter (f_0). A general approach to the problem involves model order selection [4]. In any event, the MLE of the parameters is an integral part of that approach. We now discuss the MLE, further details are available in [4].

The approximate MLE of the fundamental frequency is found by minimizing (18*b*), i.e. minimizing

$$J''(\mathbf{f}) = \mathbf{x}^H \left[\mathbf{P}_H^\perp - \mathbf{P}_H^\perp \mathbf{E} \left(\mathbf{E}^H \mathbf{P}_H^\perp \mathbf{E}\right)^{-1} \mathbf{E}^H \mathbf{P}_H^\perp\right] \mathbf{x} \ , \tag{32}$$

where $\mathbf{f} = [f_0 \, 2f_0 \cdots sf_0]^T$, or by maximizing

$$\zeta(f_0) = \mathbf{x}^H \mathbf{P}_H^\perp \mathbf{E} \left(\mathbf{E}^H \mathbf{P}_H^\perp \mathbf{E}\right)^{-1} \mathbf{E}^H \mathbf{P}_H^\perp \mathbf{x} \ . \tag{33}$$

At low SNR, an interpretation of (33) is to whiten the data (via $\mathbf{P}_H^\perp$) followed by summing the energies at the harmonics. This is a prewhitener / comb filter / energy detector. The argument is simiiar to that in [12, section 3].

In the case of a periodic signal in *white* noise, we would have to maximize

$$\zeta'(f_0) = \mathbf{x}_o^H \mathbf{E} \left(\mathbf{E}^H \mathbf{E}\right)^{-1} \mathbf{E}^H \mathbf{x}_o \ , \tag{34}$$

where $\mathbf{E} = [\mathbf{e}_1 \, \mathbf{e}_2 \cdots \mathbf{e}_s]$, and $\mathbf{e}_i = \left[1 \, e^{j2\pi i f_0} \, e^{j2\pi i f_0 2} \ldots e^{j2\pi i f_0 (N-1)}\right]^T$. If the data record is large enough so that the harmonic components are orthogonal, this becomes

$$\zeta'(f_0) \approx \mathbf{x}_o^H \mathbf{E} \, (N\mathbf{I})^{-1} \, \mathbf{E}^H \mathbf{x}_o \ , \tag{35}$$

so that

$$\zeta'(f_0) \approx \frac{1}{N} \sum_{i=1}^{s} | \, X(if_0) \, |^2 \ , \tag{36}$$

which is recognized as a comb filter / energy detector. For the colored noise case, however, we must maximize (33) over f_0, a simple one dimensional search.

We now illustrate the procedure with a simple numerical example using $N = 100$ data samples. A periodic signal with fundamental $f_0 = 0.09$ and amplitudes $A_1 = 0.5$, $A_2 = 1.0$, $A_3 = 0.5$, $A_4 = 0.5$ (with $s = 4$) is considered. Colored AR(1) noise with a pole at $0.95 \, e^{j2\pi 0.05}$ and $\sigma^2 = 1.0$ was added to the signal to generate the observed data. The fundamental frequency was estimated by minimizing the function given in (33) (the proposed method) and (36) (the usual comb filter + energy detector) using a line search over $[-0.125, 0.125)$. Note that we must have $-0.5 \leq sf_0 < 0.5$. The signal amplitudes are then computed using (20a) and the estimated fundamental frequencies to reconstruct the orignal signal. The real part of the original signal and the reconstructed versions are shown in Figure 5. The marked improvement of the proposed method over the comb filter + energy detector is mainly due to accurate estimation of the fundamental frequency. This happens because a simple comb filter + energy detector picks up the noise power and estimates the fundamental to be near the noise peak (or its subharmonic), unlike the proposed method which accounts for the noise coloration by appropriate prewhitening.

6 Conclusions

Many important problems in statistical signal processing involve separating an observed time series into a deterministic (signal) component and a random (noise) component. By modeling the noise as an autoregressive process, and the signal as either complex exponentials, polynomials, or polynomials-times-exponentials, a large class of problems can be addressed.

The main utility is due to the enormous reduction in estimator complexity, so that practical solutions to a large class of problems is now possible. Some of these problems are discussed in this paper. However, many questions such as estimation of signal modes, model order selection, etc. are yet to be investigated.

Acknowledgements: The authors would like to thank Dr. L. Pakula of the Dept. of Mathematics, Univ. of Rhode Island, Kingston, for help with the proof of Theorem 2, and Dr. P. Djuric of the Dept. of Electrical Engineering, State Univ. of New York, Stony Brook, for useful discussions.

References

[1] T. W. Anderson, *The Statistical Analysis of Time Series*. New York, NY: John Wiley, 1971.

[2] C. Chatterjee, R. L. Kashyap and G. K. Boray, "Estimation of Close Sinusoids in Colored Noise and Model Discrimination," *IEEE Trans. Acoust., Speech, Signal Process.*, pp. 328–337, Mar. 1987.

[3] D. Cochrane and G. H. Orcutt, "Application of least squares regression to relationships containing autocorrelated error terms," *J. Amer. Stat. Assoc.*, pp. 32–61, Mar. 1949.

[4] P. M. Djuric, "Selection of Signal and System Models by Bayesian Predictive Densities," Ph.D. diss., University of Rhode Island, Kingston, RI, July 1990.

[5] J. L. Doob, *Stochastic Processes*. New York, NY: John Wiley, 1953.

[6] F. R. Gantmacher, *The Theory of Matrices*. New York, NY: Chelsea Publ., 1977.

[7] I. Gohberg, P. Lancaster and L. Rodman, *Invariant Subspaces of Matrices with Applications*. New York, NY: John Wiley, 1986.

[8] N.R. Goodman, "Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution," *Ann. Math. Stat.*, pp. 152–177, 1963.

[9] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*. Stockholm, Sweden: Almqvist and Wiksell, 1956.

[10] S. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[11] S. Kay, "Asymptotically Optimal Detection in Incompletely Characterized Non-Gaussian Noise," *IEEE Trans. Acoust., Speech, Signal Process.*, pp. 627–633, May 1989.

[12] S. Kay, V. Nagesha and J. Salisbury, "Broadband Detection Based on Two-Dimensional Mixed Autoregressive Models," to appear in *IEEE Trans. Signal Process.*

[13] L. H. Koopmanns, *The Spectral Analysis of Time Series*. New York, NY: Academic, 1974.

[14] J. T. Lo, "Maximum likelihood retrieval of harmonics in colored noise by cyclic regression," *Proc. 1990 IEEE OCEANS Conf.*, pp. 369–371.

[15] R. E. Park and B. M. Mitchell, "Estimating the autocorrelated error model with trended data," *J. Econometrics*, pp. 185–201, 1980.

[16] M. B. Priestley, *Spectral Analysis and Time Series*, vol. 2. New York, NY: Academic, 1981.

[17] T.F. Quatieri and R.G. Danisewicz, "An Approach to Co-Channel Talker Interference Suppression using a Sinusoidal Model for Speech," *IEEE Trans. Acoust., Speech, Signal Process.*, pp. 56–69, Jan. 1990.

[18] D. C. Rife and R. R. Boorstyn, "Multiple Tone Parameter Estimation from discrete-time Observations," *Bell Syst. Tech. J.*, pp. 1389–1410, Nov. 1976.

[19] E. A. Robinson, *Time Series Analysis and Applications*. Houston, TX: Goose Pond Press, 1981.

[20] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*. New York, NY: John Wiley, 1989.

[21] D. Sengupta and S. Kay, "Parameter Estimation and GLRT Detection in Colored Non-Gaussian Autoregressive Processes," *IEEE Trans. Acoust., Speech, Signal Process.*, pp. 1661–1676, Oct. 1990.

[22] H. Tsuji and A. Sano, "Separable estimation of discrete and continous spectra of signal with mixed spectrum," *Proc. 1989 Int. Conf. Acoust., Speech, Signal Process.*, pp. 2218–2221.

[23] H. Wold, *A Study in the Analysis of Stationary Time Series.* Stockholm, Sweden: Almqvist and Wiksell, 1954.

[24] A. Zeira and A. Nehorai, "Frequency Domain Cramer-Rao Bound for Gaussian Processes," *IEEE Trans. Acoust., Speech, Signal Process.*, pp. 1063–1066, June 1990.

Appendix A

Equivalent Forms for Minimum Least Squares Error

We consider the generic least squares (LS) problem of minimizing

$$\mathcal{J}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = (\mathbf{x} - \mathbf{H}_1\boldsymbol{\theta}_1 - \mathbf{H}_2\boldsymbol{\theta}_2)^H (\mathbf{x} - \mathbf{H}_1\boldsymbol{\theta}_1 - \mathbf{H}_2\boldsymbol{\theta}_2) \ . \tag{A − 1}$$

The solution is easily obtained by letting $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_1^T \ \boldsymbol{\theta}_2^T\right]^T$ and $\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2]$ so that

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{x} \ , \tag{A − 2}$$

and the minimum least squares error is

$$\mathcal{J}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) = \mathbf{x}^H\mathbf{P}_H^{\perp}\mathbf{x} \ , \tag{A − 3}$$

where $\mathbf{P}_H^{\perp} = \mathbf{I} - \mathbf{H}\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H$ is the orthogonal complement projection operator.

This may also be obtained by first projecting $\mathbf{x}$ onto the subspace spanned by the columns of $\mathbf{H}_1$, and then projecting the residual onto the subspace spanned by the residual columns of $\mathbf{H}_2$.

In other words, minimizing $\mathcal{J}$ in (A-1) first with respect to $\boldsymbol{\theta}_1$ we get

$$\hat{\boldsymbol{\theta}}_1(\boldsymbol{\theta}_2) = (\mathbf{H}_1^H\mathbf{H}_1)^{-1}\mathbf{H}_1^H(\mathbf{x} - \mathbf{H}_2\boldsymbol{\theta}_2) \ , \tag{A − 4}$$

and substituting into (A-1) produces

$$\mathcal{J}(\hat{\boldsymbol{\theta}}_1(\boldsymbol{\theta}_2), \boldsymbol{\theta}_2) = \left(\mathbf{P}_1^{\perp}\mathbf{x} - \mathbf{P}_1^{\perp}\mathbf{H}_2\boldsymbol{\theta}_2\right)^H \left(\mathbf{P}_1^{\perp}\mathbf{x} - \mathbf{P}_1^{\perp}\mathbf{H}_2\boldsymbol{\theta}_2\right) \ , \tag{A − 5}$$

where $\mathbf{P}_1^{\perp} = \mathbf{I} - \mathbf{H}_1\left(\mathbf{H}_1^H\mathbf{H}_1\right)^{-1}\mathbf{H}_1^H$. Next minimizing $\mathcal{J}$ in (A-5) with respect to $\boldsymbol{\theta}_2$ we get

$$\hat{\boldsymbol{\theta}}_2 = (\mathbf{H}_2^H\mathbf{P}_1^{\perp}\mathbf{H}_2)^{-1}\mathbf{H}_2^H\mathbf{P}_1^{\perp}\mathbf{x} \ , \tag{A − 6}$$

and

$$\mathcal{J}(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) = \mathbf{x}^H\left[\mathbf{P}_1^{\perp} - \mathbf{P}_1^{\perp}\mathbf{H}_2\left(\mathbf{H}_2^H\mathbf{P}_1^{\perp}\mathbf{H}_2\right)^{-1}\mathbf{H}_2^H\mathbf{P}_1^{\perp}\right]\mathbf{x} \ , \tag{A − 7}$$

where the fact that $\mathbf{P}_1^{\perp}$ is idempotent, i.e. $\mathbf{P}_1^{\perp H} = \mathbf{P}_1^{\perp} = \mathbf{P}_1^{\perp}\mathbf{P}_1^{\perp}$ was used. The overall projection operator is $\mathbf{P}_H^{\perp} = \mathbf{P}_1^{\perp} - \mathbf{P}_1^{\perp}\mathbf{H}_2\left(\mathbf{H}_2^H\mathbf{P}_1^{\perp}\mathbf{H}_2\right)^{-1}\mathbf{H}_2^H\mathbf{P}_1^{\perp}$ which projects a vector onto the orthogonal complement of the subspace spanned by the columns of $[\mathbf{H}_1 \ \mathbf{H}_2]$. Finally, using (A-6) in (A-4), we get the overall solution for $\boldsymbol{\theta}_1$ as

$$\hat{\boldsymbol{\theta}}_1 = (\mathbf{H}_1^H\mathbf{H}_1)^{-1}\mathbf{H}_1^H\left(\mathbf{x} - \mathbf{H}_2\hat{\boldsymbol{\theta}}_2\right) \ . \tag{A − 8}$$

These expressions. (A-7) and (A-6), (A-8), were used to minimize (17) in section 3 to obtain expressions (18) and (19), (20).

Appendix B

Alternative Form for MLE of Single Frequency

Let $\hat{\mathbf{u}} = \mathbf{x} + \mathbf{H}\hat{\mathbf{a}}$, denote the estimated driving noise sequence, where $\hat{\mathbf{a}}$ is the estimate of the AR parameters which accounts for the presence of the signal. Also let $\hat{\mathbf{u}}_0 = \mathbf{P}_H^{\perp}\mathbf{x}$, denote another estimate of the driving noise wherein the AR parameters are estimated *without* accounting for the presence of the signal (i.e., using usual covariance method of AR parameter estimation). Hence, using (20a) in (20b), we obtain

$$\begin{aligned}
\hat{\mathbf{u}} &= \mathbf{x} + \mathbf{H}\hat{\mathbf{a}} \\
&= \mathbf{x} - \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\left[\mathbf{x} - \mathbf{e}\left(\mathbf{e}^H\mathbf{P}_H^{\perp}\mathbf{e}\right)^{-1}\mathbf{e}^H\mathbf{P}_H^{\perp}\mathbf{x}\right] \\
&= \mathbf{x} + \mathbf{H}\hat{\mathbf{a}}_0 + \frac{\mathbf{P}_H\mathbf{e}\,\mathbf{e}^H\mathbf{P}_H^{\perp}\mathbf{x}}{\mathbf{e}^H\mathbf{P}_H^{\perp}\mathbf{e}} \\
&= \hat{\mathbf{u}}_0 + \frac{\mathbf{P}_H\mathbf{e}\,\mathbf{e}^H\hat{\mathbf{u}}_0}{\mathbf{e}^H\mathbf{P}_H^{\perp}\mathbf{e}}
\end{aligned}$$

and thus

$$\begin{aligned}
\mathbf{e}^H\hat{\mathbf{u}} &= \mathbf{e}^H\hat{\mathbf{u}}_0 + \frac{\mathbf{e}^H\mathbf{P}_H\mathbf{e}\,\mathbf{e}^H\hat{\mathbf{u}}_0}{\mathbf{e}^H\mathbf{P}_H^{\perp}\mathbf{e}} \\
&= \frac{\mathbf{e}^H\mathbf{e}\,\mathbf{e}^H\hat{\mathbf{u}}_0}{\mathbf{e}^H\mathbf{P}_H^{\perp}\mathbf{e}}\ .
\end{aligned}$$

Using this in (21), we have

$$\begin{aligned}
\zeta(f) &= \frac{|\,\mathbf{e}^H\mathbf{P}_H^{\perp}\mathbf{x}\,|^2}{\mathbf{e}^H\mathbf{P}_H^{\perp}\mathbf{e}} \\
&= \frac{|\,\mathbf{e}^H\hat{\mathbf{u}}_0\,|^2}{\mathbf{e}^H\mathbf{P}_H^{\perp}\mathbf{e}} \\
&= \left(\mathbf{e}^H\hat{\mathbf{u}}_0\right)^*\frac{\mathbf{e}^H\hat{\mathbf{u}}_0}{\mathbf{e}^H\mathbf{P}_H^{\perp}\mathbf{e}} \\
&= \frac{\left(\mathbf{e}^H\hat{\mathbf{u}}_0\right)^*\mathbf{e}^H\hat{\mathbf{u}}}{\mathbf{e}^H\mathbf{e}} \\
&= \frac{\mathbf{e}^H\hat{\mathbf{u}}\hat{\mathbf{u}}_0^H\mathbf{e}}{\mathbf{e}^H\mathbf{e}}\ .
\end{aligned}$$

Appendix C

Invariance Proof

Let $s_1[n], s_2[n], \ldots, s_M[n]$ be a basis for $\mathcal{S}$, a subspace in $\mathcal{C}_N$. Then by definition of the invariant subspace if $x[n]$ given by

$$x[n] = \sum_{i=1}^{M} \alpha_i s_i[n] = \boldsymbol{\alpha}^T \mathbf{s}[n], \text{ for } n = 0, 1, \ldots, (N-1) ,$$

$\in \mathcal{S}$, then $y[n] = x[n] + \sum_{k=1}^{p} a[k]x[n-k] \in \mathcal{P}(\mathcal{S})$ if there exist $\beta_1, \beta_2, \ldots, \beta_M$ such that

$$y[n] = \sum_{i=1}^{M} \beta_i s_i[n] = \boldsymbol{\beta}^T \mathbf{s}[n], \text{ for } n = p, p+1, \ldots, (N-1) .$$

Thus by definition, for any $L \in \mathcal{L}$ and $x[n] = \boldsymbol{\alpha}^T \mathbf{s}[n] \in \mathcal{S}$, so that

$$y[n] = \boldsymbol{\alpha}^T \mathbf{s}[n] + \sum_{k=1}^{p} a[k] \boldsymbol{\alpha}^T \mathbf{s}[n-k]$$

we must have that $y[n] \in \mathcal{P}(\mathcal{S})$. If we choose an $L \in \mathcal{L}$ such that $a[1] = a[2] = \cdots = a[p-1] = 0$, then

$$y[n] = \boldsymbol{\alpha}^T \mathbf{s}[n] + a[p] \boldsymbol{\alpha}^T \mathbf{s}[n-p]$$

and $y[n] \in \mathcal{P}(\mathcal{S})$ iff

$$\mathbf{s}[n-p] = \mathbf{A}_p \mathbf{s}[n] \tag{C-1}$$

for some $M \times M$ matrix $\mathbf{A}_p$. Hence, we get

$$y[n] = \boldsymbol{\alpha}^T \left(\mathbf{I} + a[p]\mathbf{A}_p\right) \mathbf{s}[n] + \sum_{k=1}^{p-1} a[k] \boldsymbol{\alpha}^T \mathbf{s}[n-k] .$$

Now choosing each of the remaining $a[i]$'s to be zero except $a[j]$, with $1 \le j \le (p-1)$ (and $a[p] \ne 0$ and $\mathbf{s}[n-p]$ satisfying (C-1)), leads to

$$\mathbf{s}[n-j] = \mathbf{A}_j \mathbf{s}[n] \text{ for } 1 \le j \le (p-1) . \tag{C-2}$$

Combining (C-1) and (C-2), we have

$$\mathbf{s}[n-j] = \mathbf{A}_j \mathbf{s}[n] \text{ for } 1 \le j \le p , \tag{C-3}$$

and for $n = p, p+1, \ldots, (N-1)$.
But these conditions are all satisfied if and only if

$$\mathbf{s}[n-1] = \mathbf{A}_1\mathbf{s}[n] \text{ for } 1 \leq n \leq (N-1) \ , \tag{C-4}$$

with $\mathbf{A}_1$ being nonsingular, as we now prove. ◇

Note that (C-3) and (C-4) are trivially equivalent for $p = 1$ so that only $p \geq 2$ are of further interest. Also only the $p = 1$ case is considered in [7, p. 200].

First, we assume (C-4) is true (if part). Hence condition (C-3) for $j = 1$ is trivially true. Rewriting (C-4) using $m = n + 1$, we get

$$\begin{aligned} \mathbf{s}[m-2] &= \mathbf{A}_1\mathbf{s}[m-1] \text{ for } 2 \leq m \leq N \\ &= \mathbf{A}_1^2\mathbf{s}[m] \text{ for } 2 \leq m \leq (N-1) \end{aligned} \tag{C-5}$$

Hence condition (C-3) is true for $j = 2$. Similarly, we get

$$\mathbf{s}[n-j] = \mathbf{A}_1^j\mathbf{s}[n] \text{ for } p \leq n \leq (N-1) \ , \tag{C-6}$$

for $j = 3, 4, \ldots, p$, i.e., condition (C-3) is true, with $\mathbf{A}_j = \mathbf{A}_1^j$.

Next, we assume (C-3) is true (only if part). Then rewriting (C-3) for $j = 1$, we have

$$\mathbf{s}[n-1] = \mathbf{A}_1\mathbf{s}[n] \text{ for } p \leq n \leq (N-1) \ , \tag{C-7}$$

so that one needs to only prove that (C-4) holds for $1 \leq n \leq (p-1)$. Considering (C-3) for $j = 2$ and $n = p$, we have

$$\mathbf{s}[p-2] = \mathbf{A}_2\mathbf{s}[p] \ .$$

But $\mathbf{A}_2 = \mathbf{A}_1^2$, as we will prove henceforth, so that

$$\mathbf{s}[p-2] = \mathbf{A}_1^2\mathbf{s}[p] = \mathbf{A}_1\left(\mathbf{A}_1\mathbf{s}[p]\right)$$

and using (C-3) for $j = 1$ and $n = p$, i.e., $\mathbf{s}[p-1] = \mathbf{A}_1\mathbf{s}[p]$, we get

$$\mathbf{s}[p-2] = \mathbf{A}_1\mathbf{s}[p-1]$$

Combining the above result with (C-7), we get

$$\mathbf{s}[n-1] = \mathbf{A}_1\mathbf{s}[n] \text{ for } (p-1) \leq n \leq (N-1) \ . \tag{C-8}$$

Continuing in the same vein using $\mathbf{A}_j = \mathbf{A}_1^j$, we obtain (C-4) for $1 \leq n \leq (N-1)$.

We now prove that (C-3) also means $\mathbf{A}_j = \mathbf{A}_1^j$, as used above. We consider the proof for $\mathbf{A}_2 = \mathbf{A}_1^2$, the other cases are similar. Rewriting (C-3) for $j = 1, 2$, we have

$$\begin{aligned} \mathbf{s}[n-1] &= \mathbf{A}_1\mathbf{s}[n] \quad \text{for } p \leq n \leq (N-1) & \text{(C-9)} \\ \mathbf{s}[n-2] &= \mathbf{A}_2\mathbf{s}[n] \quad \text{for } p \leq n \leq (N-1) & \text{(C-10)} \end{aligned}$$

and letting $m = n + 1$ in (C-9) we have

$$\begin{aligned} \mathbf{s}[m-2] &= \mathbf{A}_1\mathbf{s}[m-1] \quad \text{for } (p+1) \leq m \leq (N-1) \\ &= \mathbf{A}_1^2\mathbf{s}[m] \quad \text{for } (p+1) \leq m \leq (N-1) & \text{(C-11)} \end{aligned}$$

after using (C-9) again. Combining (C-10) and (C-11), we get

$$\left(\mathbf{A}_1^2 - \mathbf{A}_2\right)\mathbf{s}[n] = \mathbf{0} \quad \text{for } (p+1) \leq n \leq (N-1) \ . \qquad (\text{C}-12)$$

This condition is satisfied in general when a vector $\mathbf{s}[n]$ lies in the null space of the $M \times M$ matrix $\mathbf{A}_1^2 - \mathbf{A}_2$. We however have M linearly independent vectors amongst $\mathbf{s}[p]$, $\mathbf{s}[p+1] \ldots$, $\mathbf{s}[N-1]$ since the columns of the matrix

$$\begin{bmatrix} \mathbf{s}^T[p] \\ \mathbf{s}^T[p+1] \\ \vdots \\ \mathbf{s}^T[N-1] \end{bmatrix} = \begin{bmatrix} s_1[p] & s_2[p] & \cdots & s_M[p] \\ s_1[p+1] & s_2[p+1] & \cdots & s_M[p+1] \\ \vdots & \vdots & \ddots & \vdots \\ s_1[N-1] & s_2[N-1] & \cdots & s_M[N-1] \end{bmatrix}$$

form a basis for $\mathcal{P}(\mathcal{S})$ (and are therefore linearly independent). In particular, we prove (by contradiction) that the last M vectors must be linearly independent, or that the vectors in $\mathcal{Q} = \{\mathbf{s}[N-M], \mathbf{s}[N-M+1], \ldots, \mathbf{s}[N-1]\}$ must be linearly independent. From (C-3),

$$\mathbf{s}[n-1] = \mathbf{A}_1\mathbf{s}[n] \quad \text{for } p \leq n \leq (N-1) \ ,$$

and using a backward recursion, we obtain

$$\begin{aligned} \mathbf{s}[N-2] &= \mathbf{A}_1\mathbf{s}[N-1] \\ \mathbf{s}[N-3] &= \mathbf{A}_1^2\mathbf{s}[N-1] \\ &\vdots \\ \mathbf{s}[N-M] &= \mathbf{A}_1^{M-1}\mathbf{s}[N-1] \ . \end{aligned}$$

Hence the set of vectors become $\mathcal{Q} = \{\mathbf{A}_1^{M-1}\mathbf{s}[N-1], \mathbf{A}_1^{M-2}\mathbf{s}[N-1], \ldots, \mathbf{A}_1\mathbf{s}[N-1], \mathbf{s}[N-1]\}$. But, by the Cayley-Hamilton theorem, any matrix satisfies its own characteristic equation or

$$A_1^M = \sum_{i=1}^{M} \alpha_i \mathbf{A}_1^{M-i} \ ,$$

so that the vectors $\mathbf{s}[N-M-1], \mathbf{s}[N-M-2], \ldots, \mathbf{s}[p]$, must all be linear combinations of the vectors in $\mathcal{Q}$. Therefore, if the vectors in $\mathcal{Q}$ are not linearly independent, then there does not exist M linearly independent vectors amongst $\mathbf{s}[p], \mathbf{s}[p+1], \ldots, \mathbf{s}[N-1]$. This contradicts the basic assumption that the sequences $s_1[n], s_2[n], \ldots, s_M[n]$ form a basis for $\mathcal{P}(\mathcal{S})$. Hence the vectors in $\mathcal{Q}$ must be linearly independent. Finally, as long as $M \leq N-p-1$, for (C-12) to be satisfied with M linearly independent vectors $\mathbf{s}[n]$, the only possibility is for the $M \times M$ matrix $\mathbf{A}_1^2 - \mathbf{A}_2$ to be the all-zero matrix, i.e., $\mathbf{A}_2 = \mathbf{A}_1^2$. The proof for $\mathbf{A}_j = \mathbf{A}_1^j$ is along the same lines except that there are $N-p-j+1$ vectors satisfying

$$\left(\mathbf{A}_1^j - \mathbf{A}_j\right)\mathbf{s}[n] = \mathbf{0} \quad \text{for } (p+j-1) \leq n \leq (N-1) \ .$$

so that we need $M \leq N-p-j+1$ for all $1 \leq j \leq p$, or that $M \leq N-2p+1$.

We finally show that the matrix $\mathbf{A}_1$ must be nonsingular. Recall that $\mathbf{s}[n]$ is a basis for $\mathcal{P}(\mathcal{S})$, i.e.,

$$\begin{aligned} \boldsymbol{\alpha}^T\mathbf{s}[n] = \mathbf{0} \quad & \text{for } p \leq n \leq (N-1) \iff \boldsymbol{\alpha} = \mathbf{0} \\ \boldsymbol{\alpha}^T\mathbf{s}[n-1] = \mathbf{0} \quad & \text{for } p+1 \leq n \leq N \iff \boldsymbol{\alpha} = \mathbf{0} \end{aligned}$$

and using (C-4) we have,

$$\begin{aligned} \boldsymbol{\alpha}^T\mathbf{A}_1\mathbf{s}[n] = \mathbf{0} \quad & \text{for } p+1 \leq n \leq (N-1) \iff \boldsymbol{\alpha} = \mathbf{0} \\ \left(\mathbf{A}_1^T\boldsymbol{\alpha}\right)^T\mathbf{s}[n] = \mathbf{0} \quad & \text{for } p+1 \leq n \leq (N-1) \iff \boldsymbol{\alpha} = \mathbf{0} \end{aligned}$$

Since there are M linearly independent vectors amongst $\mathbf{s}[p+1], \mathbf{s}[p+2], \ldots, \mathbf{s}[N-1]$, we have

$$\mathbf{A}_1^T\boldsymbol{\alpha} = \mathbf{0} \iff \boldsymbol{\alpha} = \mathbf{0}$$

i.e., $\mathbf{A}_1$ is nonsingular. ◇

We now denote $\mathbf{D} = \mathbf{A}_1^{-1}$ so that (C-4) can be written as

$$\mathbf{s}[n] = \mathbf{A}_1^{-1}\mathbf{s}[n-1] = \mathbf{D}\mathbf{s}[n-1] \quad \text{for } 1 \leq n \leq (N-1) \qquad (C-13)$$

where $\mathbf{D}$ is nonsingular. Solving the above recursion, we get

$$\mathbf{s}[n] = \mathbf{D}^n\mathbf{s}[0] \ , \qquad (C-14)$$

for $n = 0, 1, \ldots, (N-1)$. We now proceed to find all basis signals $s_i[n]$ (that comprise $\mathbf{s}[n]$) which satisfy (C-14).

Since $\mathbf{D}$ is arbitrary (only nonsingular), we consider its Jordan canonical form, i.e.,

$$\mathbf{D} = \mathbf{QJQ}^{-1}$$

where

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 & & & \mathbf{0} \\ & \mathbf{J}_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{J}_r \end{bmatrix}$$

and $\mathbf{Q}$ is a $M \times M$ modal matrix. The $\mathbf{J}_i$'s are $m_i \times m_i$ matrices, called the Jordan blocks, which are of the form

$$\mathbf{J}_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix}$$

and λ_i is a non-zero (because $\mathbf{D}$ is nonsingular) complex number, and $M = \sum_{i=1}^{r} m_i$. It is easily seen that $\mathbf{D}^n = \mathbf{QJ}^n\mathbf{Q}^{-1}$ so that (C-14) becomes

$$\mathbf{s}[n] = \mathbf{QJ}^n\mathbf{Q}^{-1}\mathbf{s}[0] \ .$$

Using $\mathbf{s}'[n] = \mathbf{Q}^{-1}\mathbf{s}[n]$ as an alternate basis for $\mathcal{C}_N$, we can rewrite the above equation as

$$\mathbf{s}'[n] = \mathbf{J}^n\mathbf{s}'[0] \ ,$$

and dropping the primes, we have

$$\mathbf{s}[n] = \begin{bmatrix} \mathbf{J}_1^n & & & \mathbf{0} \\ & \mathbf{J}_2^n & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{J}_r^n \end{bmatrix} \mathbf{s}[0] \ ,$$

and since $\mathbf{J}^n$ is block diagonal, it is enough to consider the form of $\mathbf{J}_i^n$. Denoting $\mathbf{V}_i$ to be a $m_i \times m_i$ matrix with ones along the upper second diagonal and zeros elsewhere, i.e.,

$$\mathbf{V}_i = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} ,$$

so that

$$\mathbf{J}_i = \lambda_i \mathbf{I} + \mathbf{V}_i$$

where $\mathbf{I}$ denotes the $m_i \times m_i$ identity matrix, and using the binomial expansion for $\mathbf{J}_i^n$ we obtain,

$$\begin{aligned} \mathbf{J}_i^n &= (\lambda_i \mathbf{I} + \mathbf{V}_i)^n \\ &= \sum_{l=0}^{n} \binom{n}{l} \lambda_i^{n-l} \mathbf{V}_i^l , \end{aligned}$$

where $\mathbf{V}_i^0 = \mathbf{I}$. The matrix $\mathbf{V}_i^l$ has ones along its $l+1$ upper diagonal and zeros elsewhere and $\mathbf{V}_i^l = \mathbf{0}$ for $l \geq m_i$, as can be easily verified. Hence,

$$\mathbf{J}_i^n = \sum_{l=0}^{m_i - 1} \binom{n}{l} \lambda_i^{n-l} \mathbf{V}_i^l ,$$

which may be explicitly written as

$$\mathbf{J}_i^n = \begin{bmatrix} \lambda_i^n & \binom{n}{1}\lambda_i^{n-1} & \binom{n}{2}\lambda_i^{n-2} & \cdots & \binom{n}{m_i - 1}\lambda_i^{n-(m_i-1)} \\ 0 & \lambda_i^n & \binom{n}{1}\lambda_i^{n-1} & \cdots & \binom{n}{m_i - 2}\lambda_i^{n-(m_i-2)} \\ 0 & 0 & \lambda_i^n & \cdots & \binom{n}{m_i - 3}\lambda_i^{n-(m_i-3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i^n \end{bmatrix} . \qquad (C-15)$$

Partitioning $\mathbf{s}[n]$ as $\left[\mathbf{s}_1^T[n]\ \mathbf{s}_2^T[n]\ \cdots\ \mathbf{s}_r^T[n]\right]^T$, where $\mathbf{s}_i[n]$ is $m_i \times 1$, the solution becomes

$$\mathbf{s}_i[n] = \mathbf{J}_i^n \mathbf{s}_i[0] \ .$$

Examining the form of $\mathbf{J}_i^n$ more closely, and denoting $\mathbf{s}_i[0] = [c_0\, c_1\, \cdots\, c_{m_i-1}]^T$, we rewrite the above equation as

$$\begin{aligned}
\mathbf{s}_i[n] &= \mathbf{J}_i^n \mathbf{s}_i[0] \\
&= \begin{bmatrix}
\lambda_i^n & \binom{n}{1}\lambda_i^{n-1} & \binom{n}{2}\lambda_i^{n-2} & \cdots & \binom{n}{m_i-1}\lambda_i^{n-(m_i-1)} \\
0 & \lambda_i^n & \binom{n}{1}\lambda_i^{n-1} & \cdots & \binom{n}{m_i-2}\lambda_i^{n-(m_i-2)} \\
0 & 0 & \lambda_i^n & \cdots & \binom{n}{m_i-3}\lambda_i^{n-(m_i-3)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \lambda_i^n
\end{bmatrix}
\begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{m_i-1} \end{bmatrix} \\
&= \begin{bmatrix}
c_0 & c_1 & c_2 & \cdots & c_{m_i-1} \\
c_1 & c_2 & c_3 & \cdots & 0 \\
c_2 & c_3 & c_4 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
c_{m_i-2} & c_{m_i-1} & 0 & \cdots & 0 \\
c_{m_i-1} & 0 & 0 & \cdots & 0
\end{bmatrix}
\begin{bmatrix}
\lambda_i^n \\ \binom{n}{1}\lambda_i^{n-1} \\ \vdots \\ \binom{n}{m_i-1}\lambda_i^{n-(m_i-1)}
\end{bmatrix}
\end{aligned} \qquad \text{(C-16)}$$

Finally, the signals $\{\lambda_i^n, \binom{n}{1}\lambda_i^{n-1}, \binom{n}{2}\lambda_i^{n-2}, \cdots, \binom{n}{m_i-1}\lambda_i^{n-(m_i-1)}\}$ can be easily shown to be linearly independent. Since we can assume without loss of generality that $c_{m_i-1} \neq 0$ (it only decides the dimensionality of the subspace) and hence the matrix is nonsingular, the elements of $s_i[n]$ will also be linearly independent and will span the same subspace as the signals $\{\lambda_i^n, \binom{n}{1}\lambda_i^{n-1}, \binom{n}{2}\lambda_i^{n-2}, \cdots, \binom{n}{m_i-1}\lambda_i^{n-(m_i-1)}\}$, which is a m_i dimensional subspace $\mathcal{S}_i$. The number of these subspaces is arbitrary as long as $\sum_{i=1}^r m_i = M$. The entire subspace spanned by these signals (with different λ_i), i.e., $\mathcal{S}$, is the direct sum of the subspaces, $\mathcal{S}_i$, or $\mathcal{S} = \mathcal{S}_1 \oplus \mathcal{S}_2 \oplus \cdots \oplus \mathcal{S}_r$. $\diamond$

Appendix D

General Parameter Transformation

To determine the general parameter transformation, consider the signal

$$s[n] = \sum_{i=1}^{r} \sum_{j=0}^{m_i-1} B_j^{(i)} \binom{n}{j} \lambda_i^{n-j} \ . \qquad \text{(D - 1)}$$

Since each subspace $\mathcal{S}_i$ is invariant, we need to consider only

$$s_i[n] = \sum_{j=0}^{m_i-1} B_j^{(i)} \binom{n}{j} \lambda_i^{n-j} \ . \qquad \text{(D - 2)}$$

Let $\mathbf{B}_i = \left[B_0^{(i)}\, B_1^{(i)} \cdots B_{m_i-1}^{(i)}\right]^T$, and since the first row of $\mathbf{J}_i^n$ contains the basis signals, we obtain

$$\begin{aligned} s_i[n] &= \mathbf{B}_i^T \mathbf{s}_i[n] \\ &= \left[B_0^{(i)}\, B_1^{(i)} \cdots B_{m_i-1}^{(i)}\right] \begin{bmatrix} \lambda_i^n \\ \binom{n}{1}\lambda_i^{n-1} \\ \binom{n}{2}\lambda_i^{n-2} \\ \vdots \\ \binom{n}{m_i-1}\lambda_i^{n-(m_i-1)} \end{bmatrix} \\ &= \mathbf{B}_i^T \mathbf{J}_i^{n\,T} \mathbf{e}_1 \ , \end{aligned}$$

where $\mathbf{e}_1 = [1\,0\,0 \cdots 0]^T$. But,

$$\begin{aligned} y_i[n] &= \sum_{k=0}^{p} a[k] s_i[n-k] \\ &= \sum_{k=0}^{p} a[k] \mathbf{B}_i^T \mathbf{J}_i^{n-k\,T} \mathbf{e}_1 \\ &= \mathbf{B}_i^T \left(\sum_{k=0}^{p} a[k] \mathbf{J}_i^{-k\,T} \right) \mathbf{J}_i^{n\,T} \mathbf{e}_1 \\ &= \mathbf{D}_i^T \mathbf{s}_i[n] \end{aligned}$$

where

$$\mathbf{D}_i = \left(\sum_{k=0}^{p} a[k] \mathbf{J}_i^{-k} \right) \mathbf{B}_i$$

for $i = 1, 2, \ldots, r$.

Figure 2a: CARD Spectral Estimate

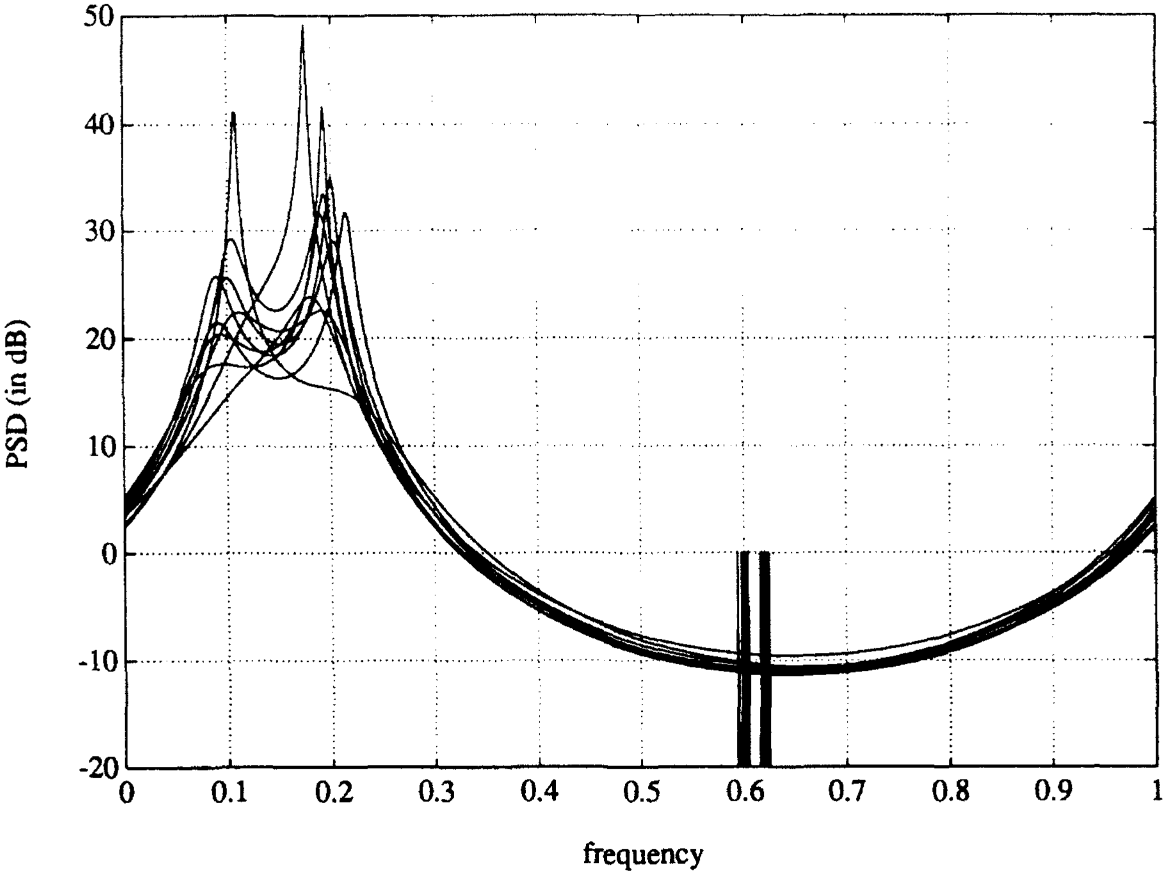


Figure 2b: Line Spectral Estimate

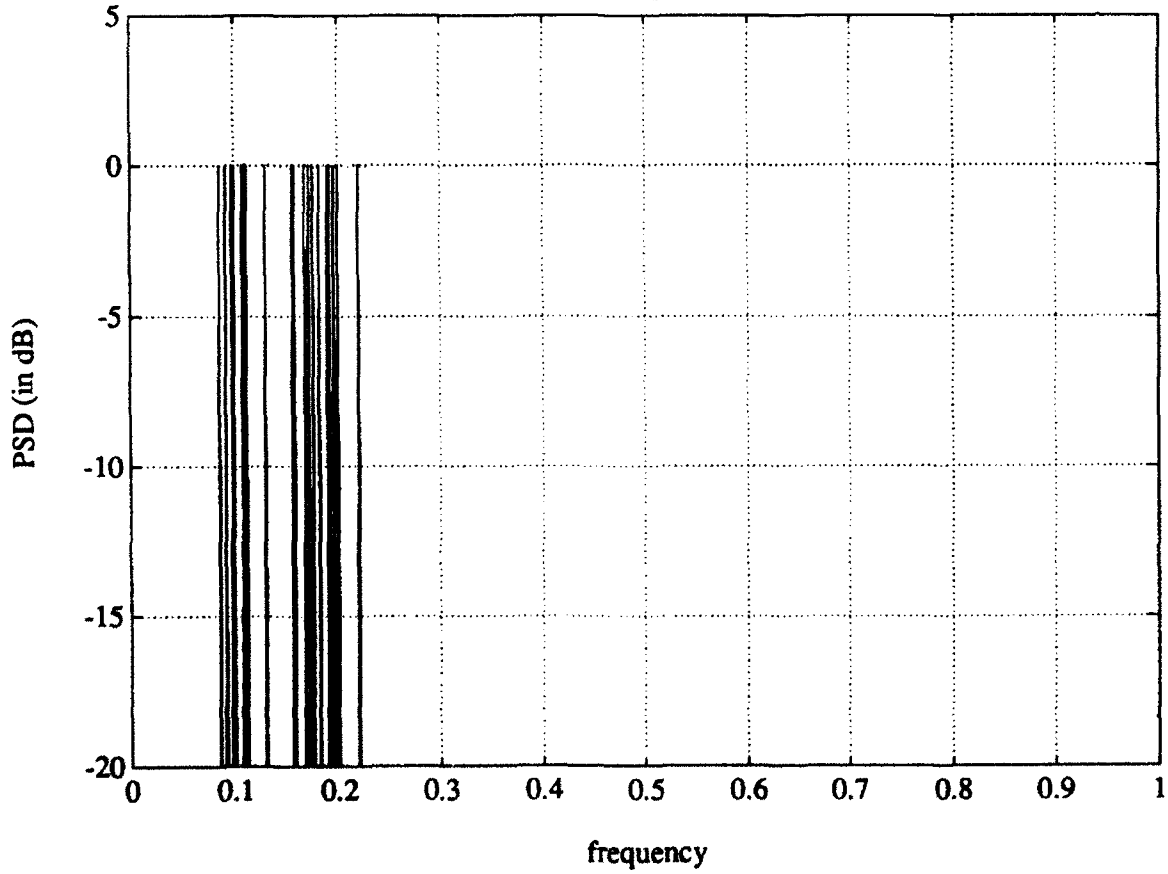


Figure 2c: AR Spectral Estimate (modified covariance method)

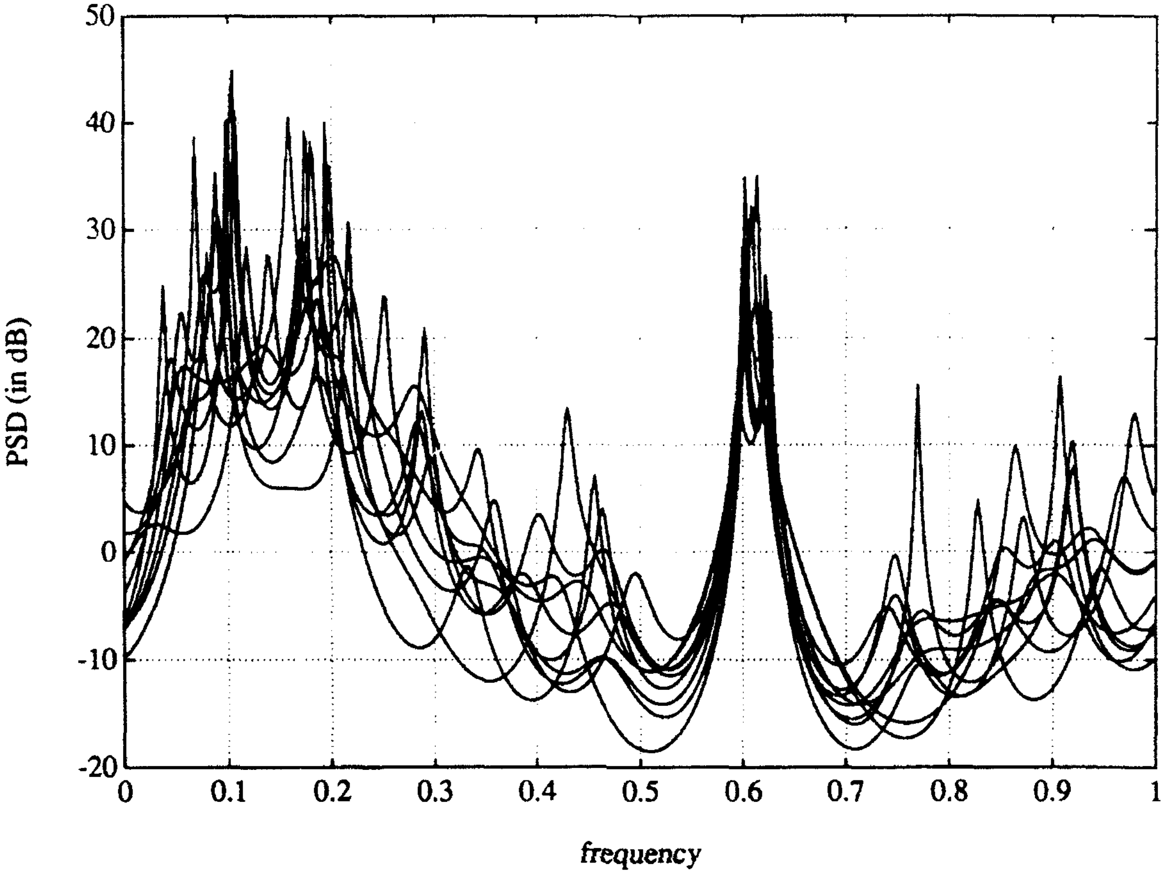


Figure 2d: Average Spectral Estimates using 50 realizations

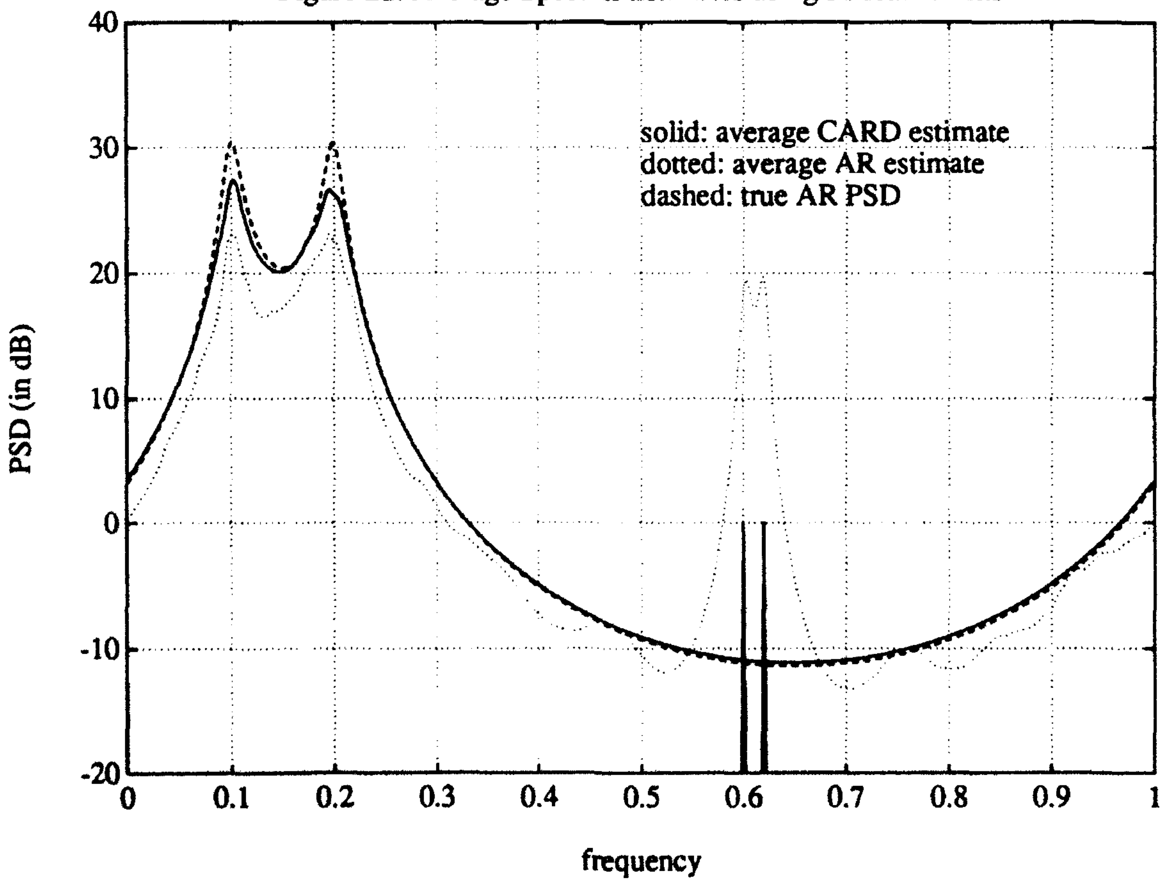


Figure 3a: CARD Spectral Estimate

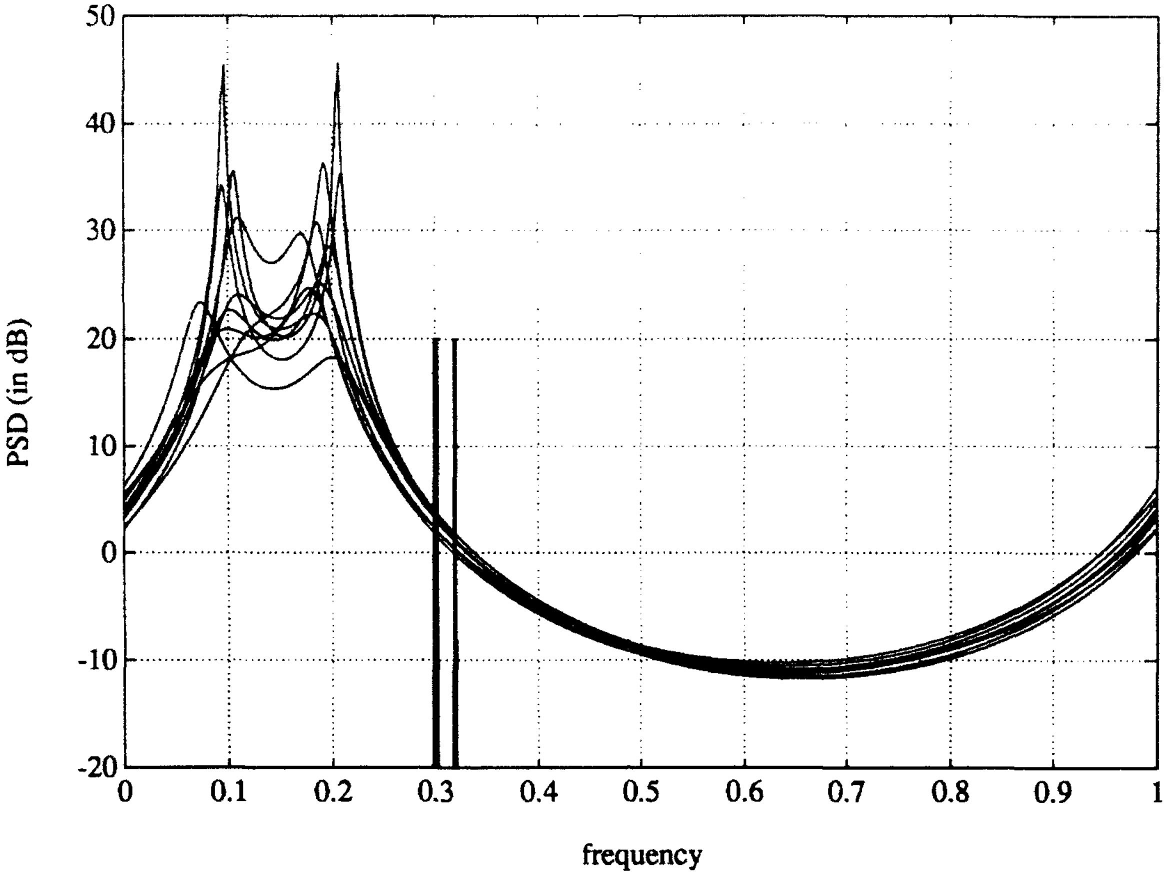


Figure 3b: Line Spectral Estimate

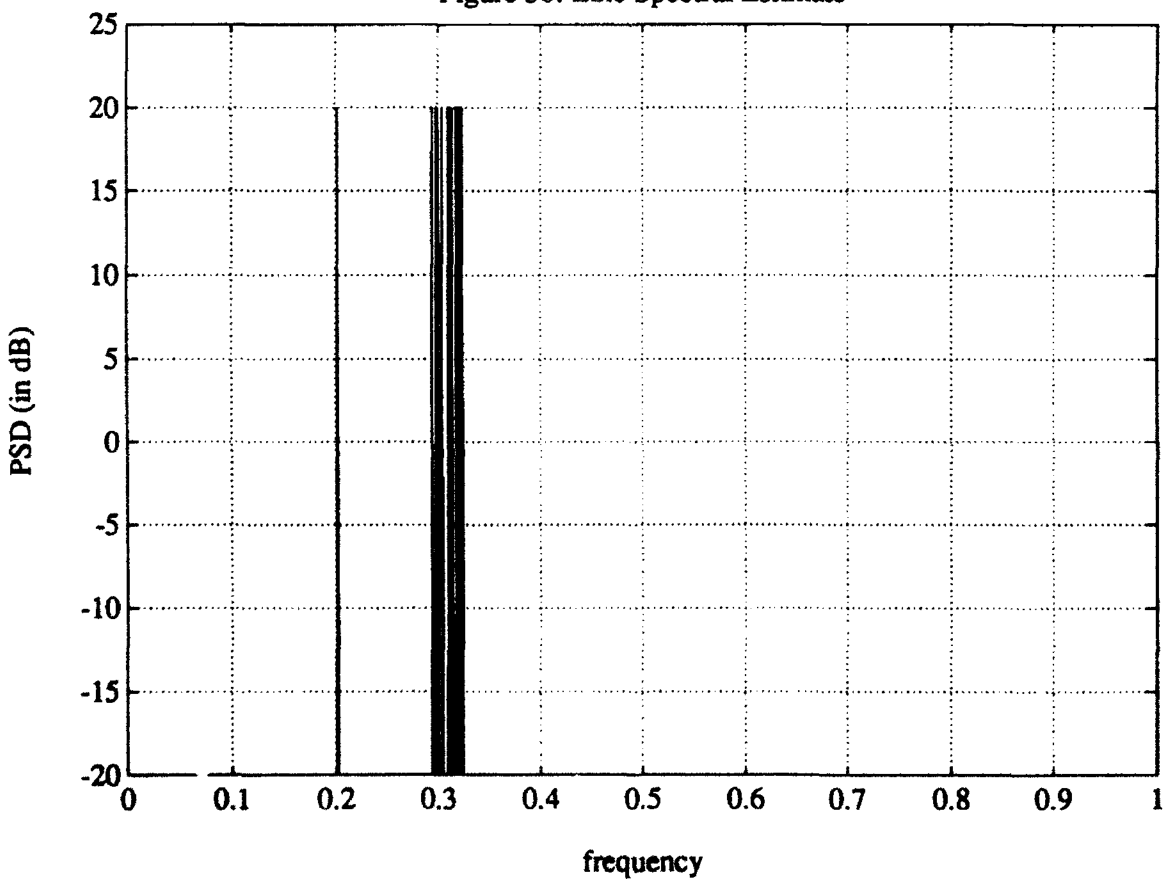


Figure 3c: AR Spectral Estimate (modified covariance method)

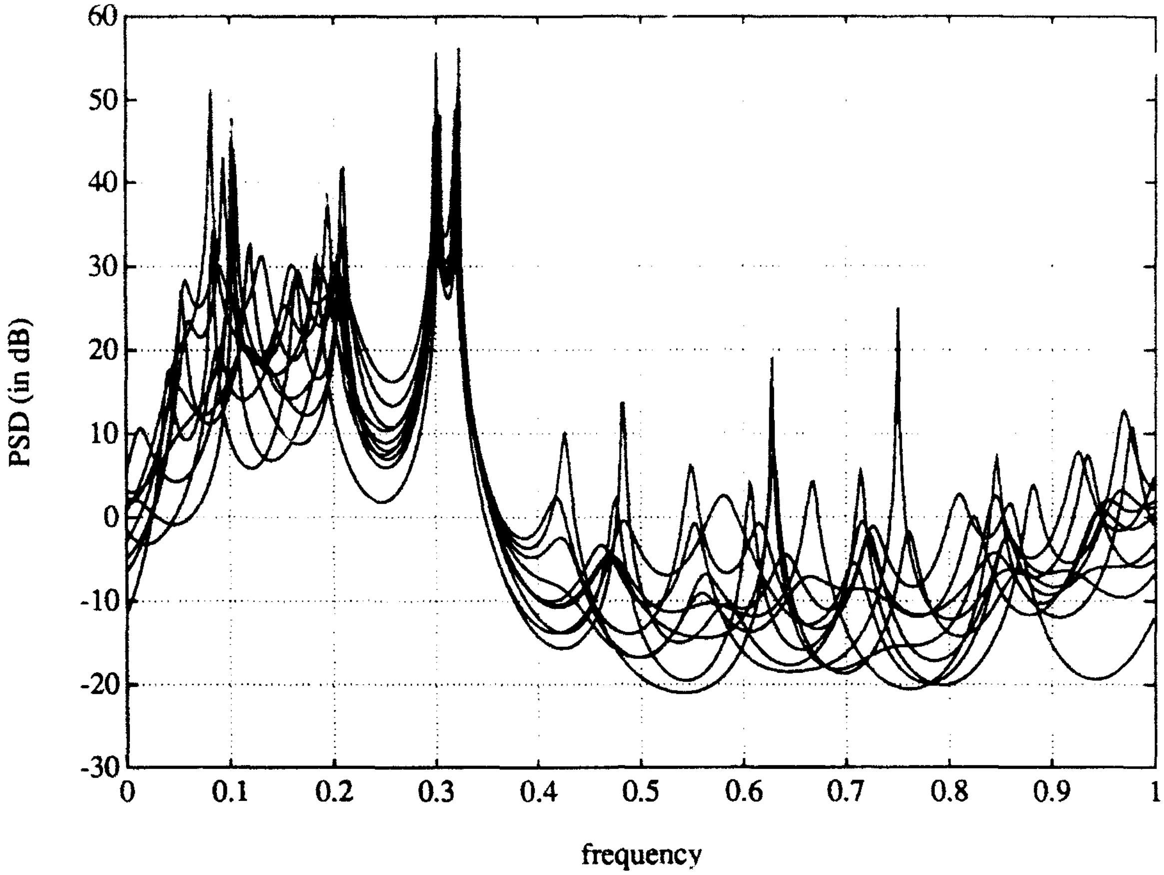


Figure 3d: Frequency Estimates using CARD (solid) and line spectra (dashed)

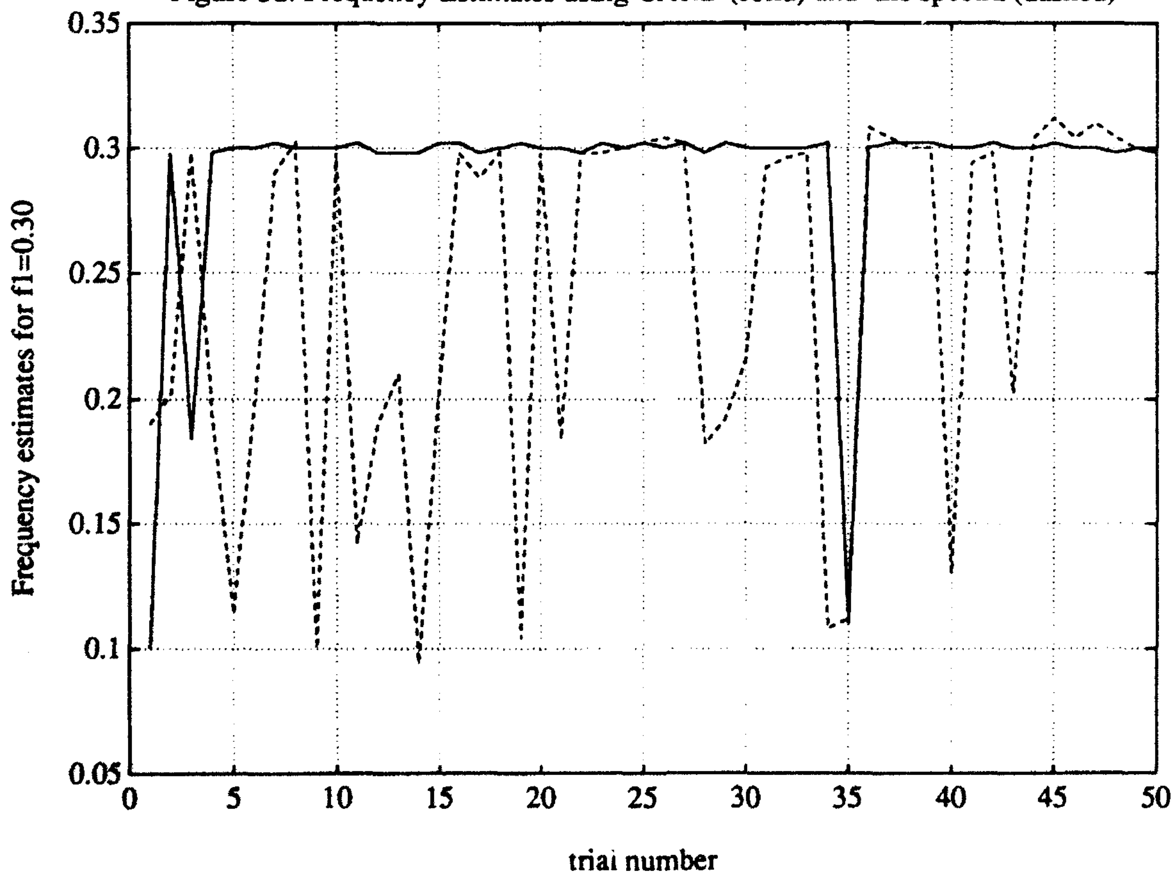


Figure 3e: Frequency Estimates using CARD (solid) and line spectra (dashed)

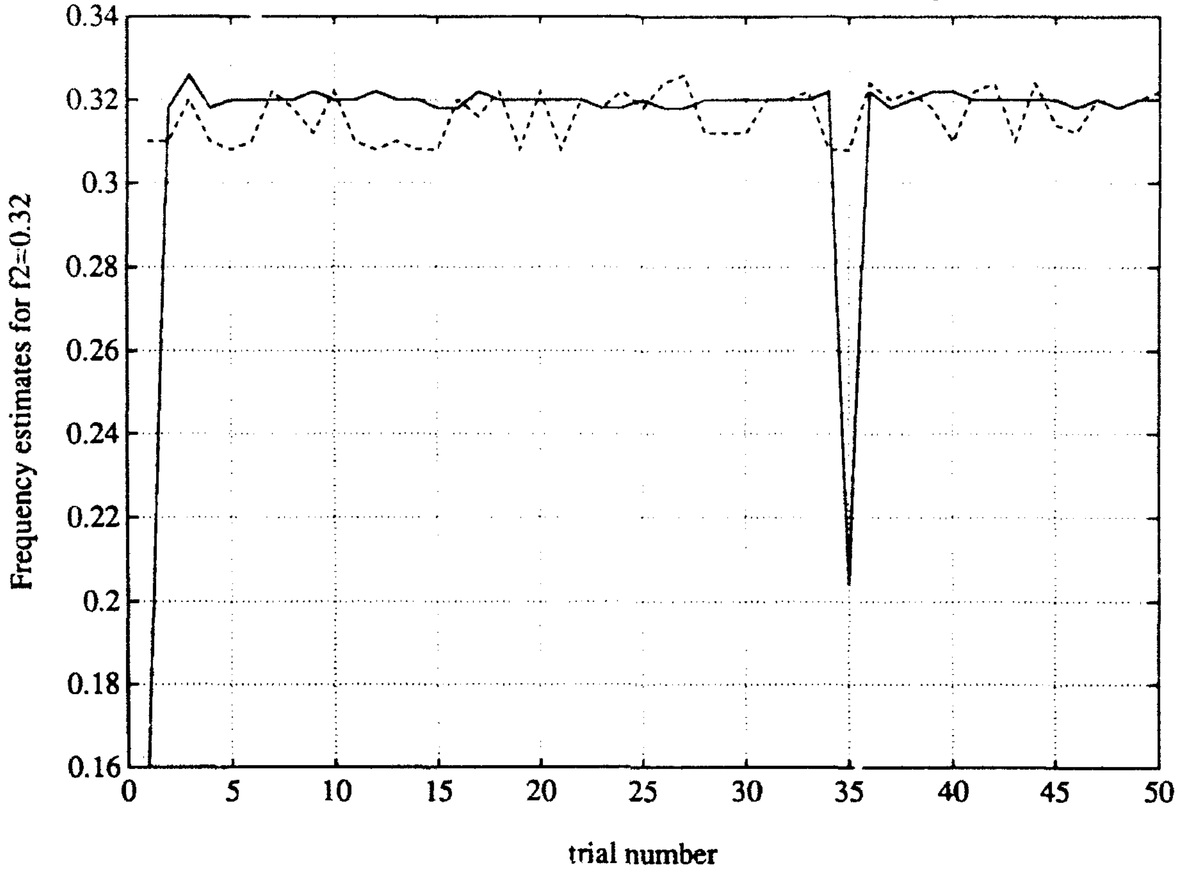


Figure 3f: Average Spectral Estimates using 50 realizations

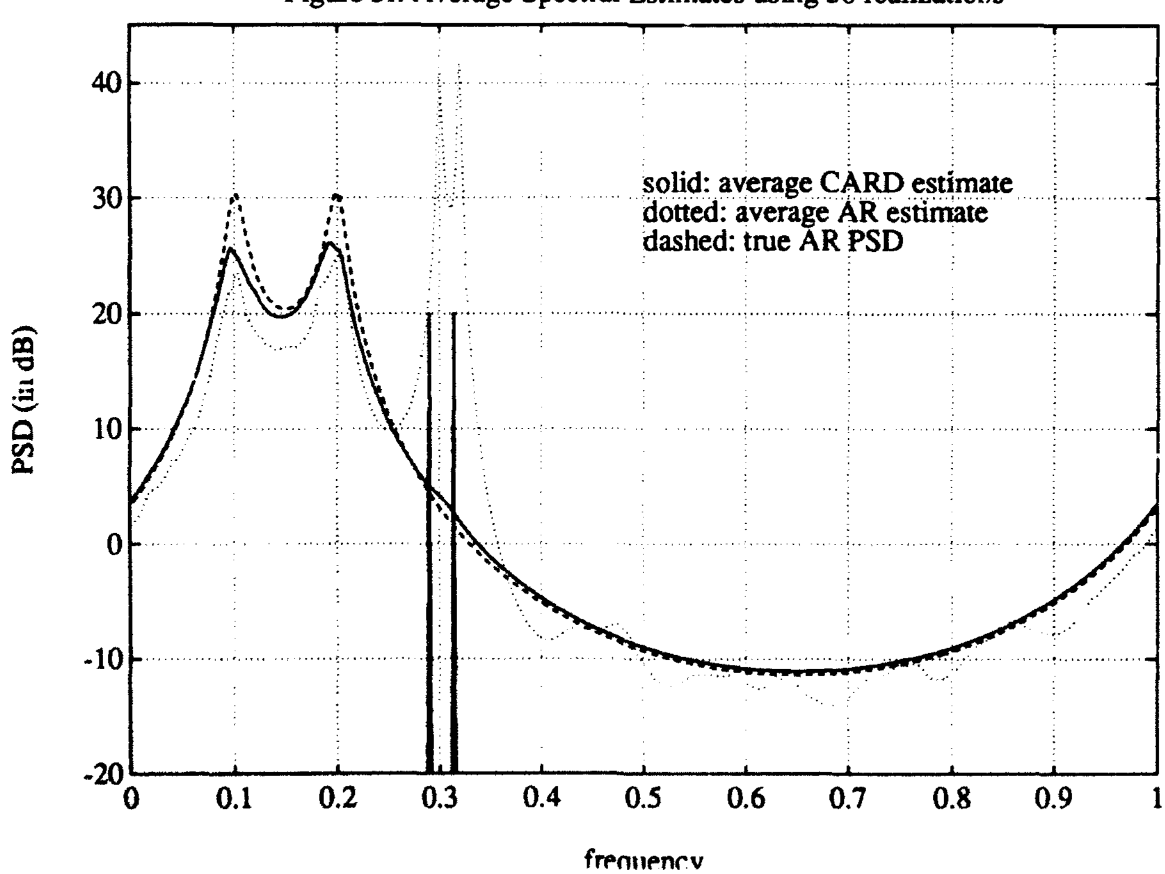


Figure 4a: CARD Spectral Estimate

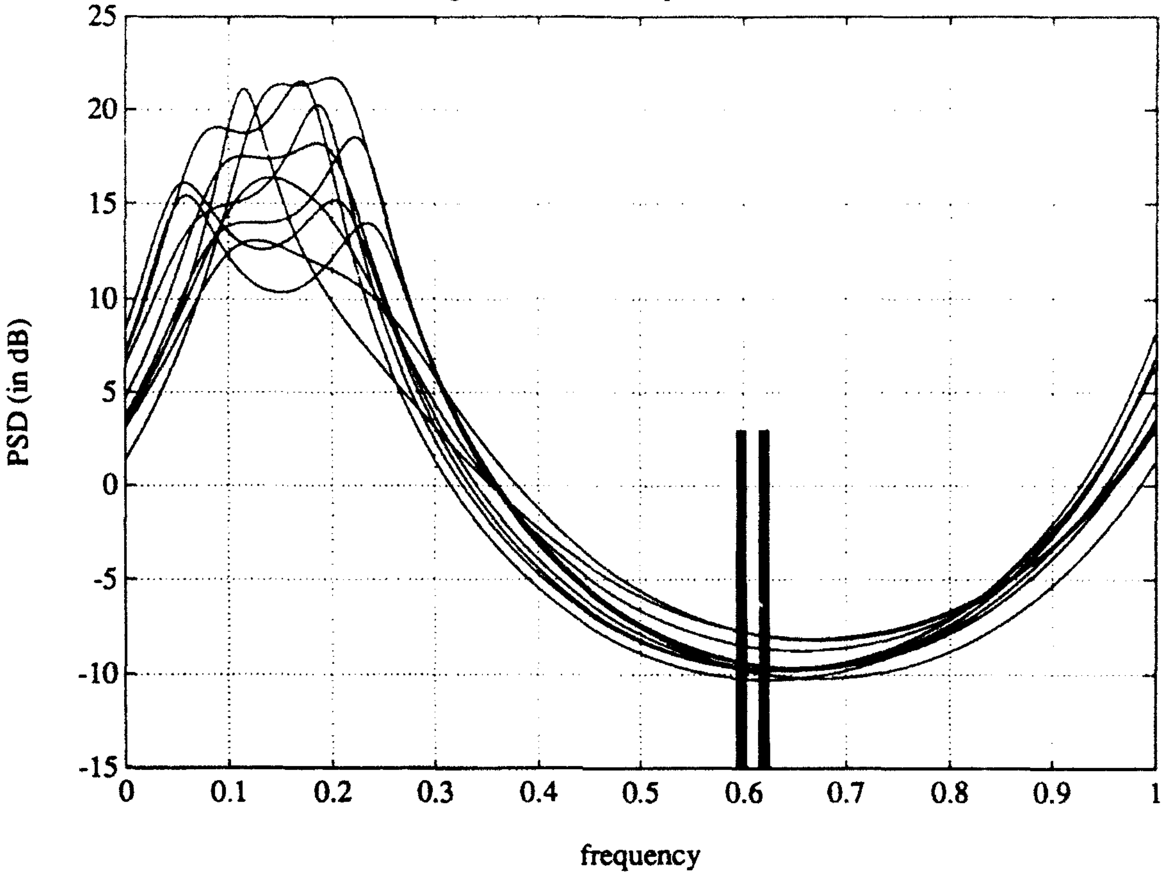


Figure 4b: Line Spectral Estimate

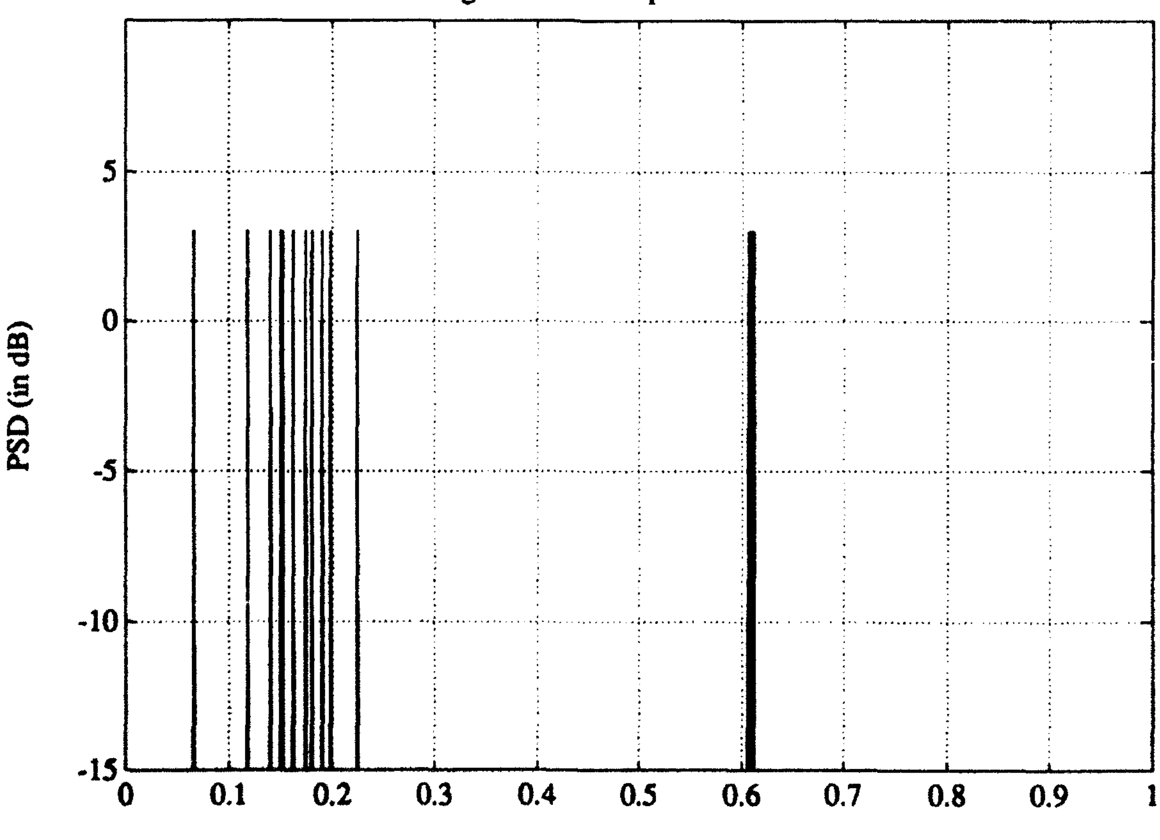


Figure 4c: AR Spectral Estimate (modified covariance method)

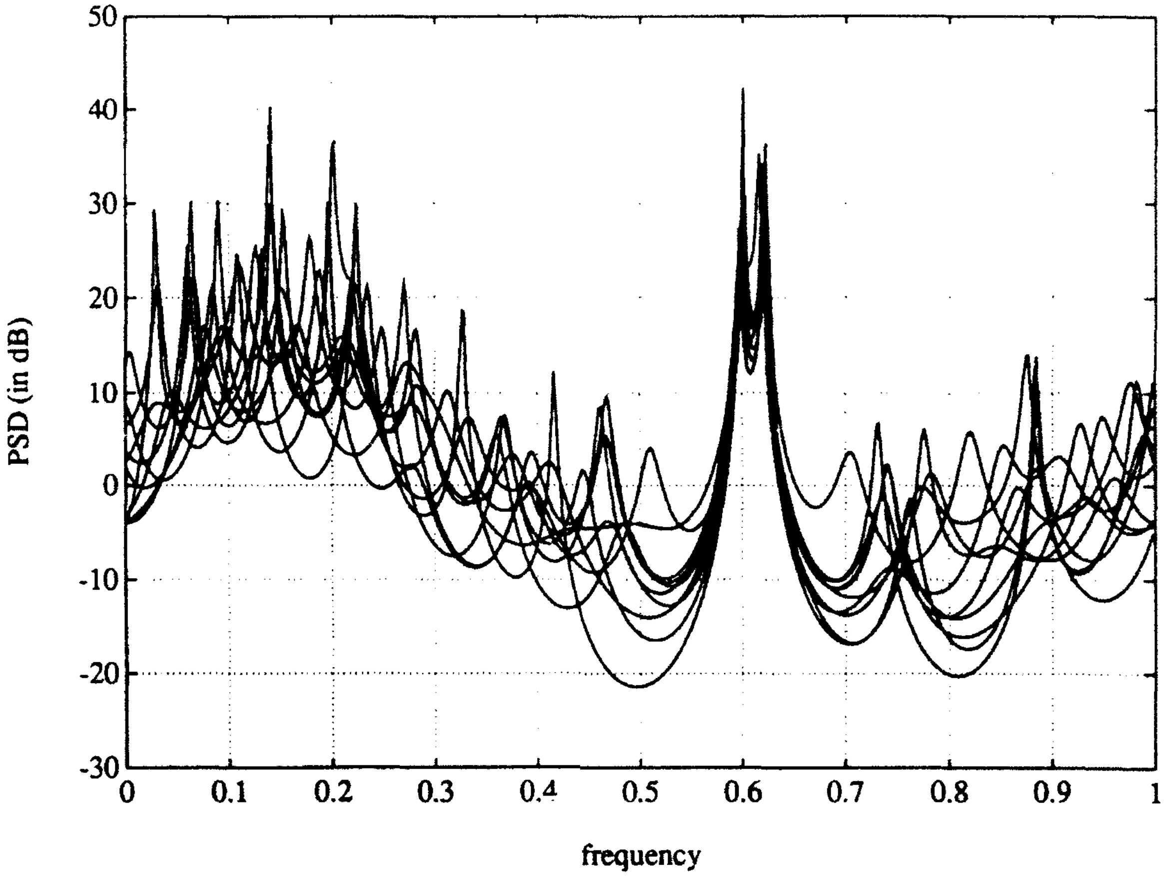


Figure 4d: Average Spectral Estimates using 50 realizations

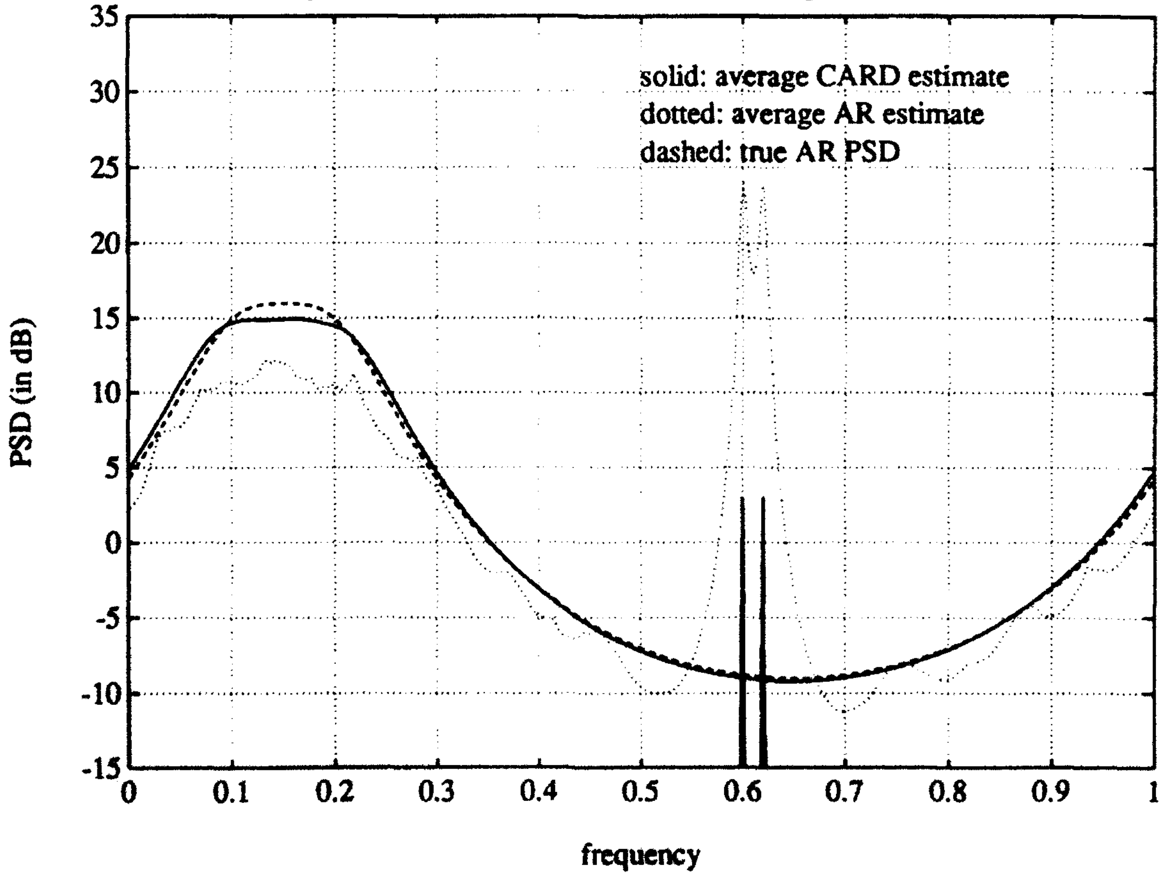


Figure 5: Reconstruction of Periodic Signal